Lenition in the production and perception of Chilean Spanish approximant consonants: Implications for lexical access models

Mauricio Alejandro Figueroa Candia

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

to
Department of Speech, Hearing and Phonetic Sciences
University College London (UCL), London, UK

2016

Declaration

I, Mauricio Alejandro Figueroa Candia, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Mauricio Alejandro Figueroa Candia

Abstract

Chilean Spanish approximant consonants [β̞ ð̞ ɣ̞] display high degrees of lenition, which often leads to elision in several phonetic contexts. The fact that these units can surface at any stage in a continuum from approximant to elided is an ideal testing ground for exploring how listeners attain lexical access while coping with varying degrees of acoustic and semantic information, and provides important evidence for evaluating the predictions that lexical access models make about processing highly degraded and variable signals.

An initial production study was conducted to determine the scope of lenition for /b d g/ in Chilean Spanish. Ten native speakers were recorded while completing three elicitation tasks: word-lists, short texts and semi-guided conversations. Several duration, intensity and formant measurements were extracted, normalized and analysed. The results showed that lenition and elided variants are indeed a common feature of these consonants, and that the relevant variability is encoded in the interaction between duration, intensity and F1.

Given this variation, the second study investigated how listeners resolve potential ambiguities in speech processing. Continua from approximant consonant to elision were prepared and presented to listeners in conditions which varied in the degree of acoustic and semantic cues available, in several perception tasks: phoneme monitoring, identification and discrimination. For phoneme monitoring and identification, the results for /b/ and /d/ showed category boundary shifts when semantic information became available, but no further semantic priming effects. No significant category boundary shifts were observed for /g/. The results from the discrimination tasks, on the other hand, showed that sensitivity to differences between consonant presence and elision rises as lenition increases.

The results from the production and perception studies are discussed in the light of lexical access models, in particular with regard to the divide between abstractionist, episodic and hybrid models.

Acknowledgements

First and foremost, I want to express my deepest gratitude to my main supervisor, Bronwen Evans, whose support and encouragement along this process have been paramount. Her ideas, insight, suggestions and guidance always greatly improved my research, while at the same time refined my points of view. I also sincerely thank her for allowing me to collaborate with her in teaching and supervising, all of which I thoroughly enjoyed.

I am also indebted to other members of the Speech, Hearing and Phonetic Sciences department, and the Linguistics department. Firstly, thanks to my second supervisor, John Harris, and to Paul Iverson and Patti Adank, for very useful feedback given to me during my upgrade. Secondly, thanks to Mark Huckvale and Stuart Rosen for their help with several methodological issues.

This dissertation was assisted in several ways by Chilean colleagues, to all of whom I sincerely say *gracias*. Thanks to Domingo Román for help collecting data for the first production study; to Camilo Quezada for early advice on statistical analyses; and to Rosa Catalán, Nicolás Retamal, Josué Pino and Claudia Castro, for assistance collecting data for perception experiments.

To all the anonymous participants, my deepest gratitude. Your names will be forgotten –pursuant UK's Data Protection Act 1998–, but your responses will remain forever plotted.

I had the pleasure to share my postgraduate experience with an amazing group of fellow-students and post-doctoral seniors. Many thanks first to José Joaquín Atria, Kurt Steinmetzger (and Natalie), Tim Schoof and Michèle Pettinato, for friendship, encouragement and for many inspired conversations. Thanks also to Georgina Oliver, Wafa Alshangiti, Nada Al-Sari and Melanie Pinet, for your friendship and generous help along my wife's and my own experience in London, and particularly while enduring our crash-course into parenthood. Thanks also to my dear friends Sonia Granlund, Yasuaki Shinohara (my Chop-Chop buddy), Alberto Lee, Gisela Tomé and Yasna Pereira.

This dissertation was funded by *Becas Chile*, from the Chilean Ministry of Education, via CONICYT. Thanks to all the corresponding bodies and agencies for financial support during these years.

Finally, many thanks to my family. First to my parents, Patricia and Dante, for continuous support. Thanks too to Isabel and Juan Manuel, my in-laws, for visiting us several times in London and for your help during very challenging months with the baby. Many thanks, finally, to Paula, *compañera de mis días y del porvenir*. Thank you so much for joining me in this rather crazy journey.

Contents

List of figures

List of tables

List of equations

Chapter 1

Summary of chapters

The aim of this thesis is threefold: firstly, to investigate the scope of lenition in Chilean Spanish approximant consonants of /b d g/; secondly, to determine how listeners deal with lenited and elided instances of said segments, in particular how listeners respond to the availability of acoustic and semantic cues; and thirdly to analyse the results obtained in the production and perception studies in the light of traditional and current models of lexical access and speech perception.

1.1. Chapter-by-chapter summary

Chapter 2 provides a general introduction to the theoretical topics relevant to the dissertation, and at the same time summarizes previous research for the study of the variation of Spanish and Chilean Spanish /b d g/[1]. Firstly, several definitions of lenition and variables conditioning it are reviewed. Then, cue weighting is described briefly, placing particular attention on the role of non-acoustic cues in perception. In the next sub-section, two main perceptual "illusions" are considered: phonological recovery, whereby listeners perceive units not present in the acoustic signal, and lexical effects on speech perception. Three main families of lexical access models are reviewed afterwards (abstractionist, episodic and hybrid), stressing their main differences and similarities. Next, the phonetic variability in Spanish and Chilean Spanish /b d g/ is reviewed in detail. The chapter finishes with a discussion relating to how this research is relevant for this dissertation.

Chapter 3 presents the methods used to collect and normalize acoustic data for approximant variants of /b d g/. A brief summary of the state of the art for Spanish and Chilean Spanish /b d g/ is provided, followed by a short discussion on methodological issues and a summary of the aims of this part of the thesis. The methods employed to obtain data from production are described in detail, covering general characteristics of

1 The place of articulation for Spanish /d/ is postdental (e.g., Sadowsky & Salamanca, 2011), and thus a strict compliance to the IPA would require to transcribe it as /d̪/. Here, /d/ is used to avoid over-complicating transcription.

participants, elicitation procedures, the technical set-up, segmentation, labelling and coding protocols, and the variables of interest. Lastly, the normalization procedures applied to the acoustic data for duration, intensity and formants are explained and evaluated, and some suggestions are made regarding the application of methods for future research.

Chapter 4 presents the results from the production study. After an overview of the results, the acoustic properties of the approximant variants of /b d g/ are presented in detail. In the case of duration, intensity and F1, their variation was found to be in line with previous accounts and theoretical assumptions. In particular, more lenited variants showed shorter duration and higher intensity and F1 values. This section concludes that duration, intensity and F1 encode the relevant variation for degree of lenition, while the role of F2 is less clear. The effect of phonetic context, word status, internal word frequency and elicitation procedure on the variability of /b d g/ is then considered, with results confirming previous research and theoretical expectations: weak phonetic contexts (e.g., intervocalic) favour lenition, and more lenition is found in words (as opposed to nonsense words), higher frequency words and elicitation tasks in which the participant pays less attention to speech. The chapter concludes with a discussion of the relevance of the findings for previous research precedents and following chapters. Two focal topics are Chilean Spanish as a particularly lenited variant from Spanish, and issues on methodological standards in production studies involving spirant approximants.

Chapter 5 describes a small perception study aimed at determining whether listeners are able to identify and discriminate bilabial and labiodental approximant consonants of /b/. The chapter begins with a short summary of the state of the art for /b/, and a synthesis of the main results found in the production study (Chapter 4). In this experiment, natural examples of [β̞] and [ʋ], and synthetic continua from [β̞] to [ʋ] were presented in identification and discrimination tasks. Statistical results showed that participants displayed some sensitivity to the bilabial versus labiodental contrast in natural approximant consonants, although considerable overlap was found between categories and there was a large amount of individual variability. No effect of stimulus level was found on the identification and discrimination responses for synthetic stimuli. Finally, the implications of these results are discussed, after which two main conclusions are offered: first, that listeners are not able to identify of discriminate [β] and [ʋ], and second, that this variation can safely be disregarded in subsequent perception experiments.

Chapter 6 describes a series of perception experiments which aimed to investigate the perception of Chilean Spanish [β̞ ð̞ ɣ̞]. An introduction is provided first, in which the literature on perception of highly lenited variants, phonological recovery and their relation to lexical access models is reviewed. Additionally, the theoretical and methodological relevance of the natural variation of Chilean Spanish /b d g/ for these issues is highlighted. Listeners completed phoneme monitoring, identification and discrimination tasks, in which synthetic continua from full approximants to elided variants (in which both ends were separate legal Spanish words) were presented in several informational conditions: from only minimal segmental cues to semantic priming of word-level stimuli. Results showed that increasing the amount of acoustic and semantic cues had an effect on listeners' responses, enabling lexical effects and phonological recovery, and bringing responses closer to categorical perception distributions. A second main finding was that these effects were relatively different for the three consonants being tested, and that these differences seemed to depend on the listeners' expectations regarding what is normal in natural production and perception. The chapter findings are discussed in light of previous research on lexical effects and recovery, with particular attention to how the results can be accounted for by major families of lexical access models. The chapter finishes by discussing some limitations of the study that might have prevented clearer semantic priming effects on the results.

Chapter 7 presents the results from a small follow-up study that aims to further explore semantic priming effects on the perception of approximant variants of /b d g/. After a brief introduction, a methods section is provided, describing the design of synthetic continua from full approximant to elided variants for /b d g/, in which the full approximant end of the continua was a (relatively) high frequency word and the elided end of each continuum could not be interpreted as a separate legal Spanish word. Continua were embedded into two informational conditions (word-level and primed word-level) and presented in phoneme monitoring tasks. Inspection of the response distributions and of the results from statistical analyses revealed that no semantic priming effects had taken place. Also, lexical effects seemed to affect [β̞] and [ɣ̞] to a greater extent when compared to [ð̞]. The discussion addresses the main findings from the experiments, attributing the failure to observe semantic priming effects to methodological limitations, and positing that the differences observed in the responses for /b d g/ are related to listeners' expectations regarding natural production and perception. As in previous chapters, the discussion relates the main findings to previous research on recovery, lexical effects on perception, and models of lexical access.

Chapter 8 provides a general discussion of the main results from this thesis, aiming to explain the implications of the experimental findings for the general literature on lexical effects on speech perception, phonological recovery and lexical access models. Two major discussion points are how lenition of /b d g/ reveals a link between production and perception, and the relative advantages of hybrid models of lexical access at accounting for our findings about perception. The general discussion finishes by addressing some general limitations and by making suggestions for future research.

Chapter 2
General introduction

This chapter introduces the theoretical areas and research precedents relevant to this thesis, and begins by considering different definitions of *lenition*, after which well known examples of the phenomenon will be provided, and several factors conditioning it will be summarized. Next, *cue weighting* will be discussed, with the emphasis on previous research investigating how contextual and semantic cues can aid the perception of lenited and elided units. The following section will be devoted to two types of "speech illusions": *phonological recovery*, in which units lacking acoustic evidence are still perceived by listeners, as in cases of highly lenited units, and *lexical effects* on perception, in which the status of a word –whether a word or nonsense word– can affect prelexical levels of speech processing (i.e., the Ganong effect).

The following subsections provide a summary of a selection of traditional and contemporary abstractionist, episodic and hybrid models of lexical access and speech perception. After these, a synthesis along the abstractionist/episodic axis will be provided in three separate subsections, including one for hybrid models. The last section of this chapter describes the approximant consonants of Spanish /b d g/. The definition of the term *approximant* will be first discussed in detail. Then, the characteristics of Spanish /b d g/ will be summarized by conducting a literature review of the most relevant studies addressing the articulatory and acoustic characteristics of approximant variants, before focussing in more detail on Chilean Spanish itself. The final subsection will provide a link between the topics developed in the introduction and Chapter 3, where the methods for a large scale production study on the approximant variants of /b d g/ are detailed.

2.1. Lenition

Arriving at a comprehensive and unified definition of lenition has proven to be quite difficult. In its most simple form, lenition –or weakening– has been understood as a reduction process that often results in deletion (Escure, 1977). Other approaches that also make reference to the directionality of changes suggest that a segment “A” can be considered to be more lenited than another segment “B” if “B” becomes “A” at some stage during a sound change process which culminates in deletion (Vennemann, 1988). Definitions like these have been challenged on several grounds. For example, it has been pointed out that directionality should not be incorporated into definitions of lenition, since some changes that are normally considered to form part of a weakening process have reversed to fortition (e.g., from [ʎ] to [j̞], but then to [d͡ʒ], in some dialects of Spanish), or that not all cases of lenition result in deletion (Bauer, 2008).

A different approach posits that lenition should be defined as a degradation of the informational complexity of the speech signal, since lenited units carry fewer acoustic cues than non-lenited parent units, thus reducing the phonological complexity of segments (Harris & Urua, 2001). A similar proposal, developed partially in response to approaches that see lenition as effort reduction (e.g., Kohler, 1990; Lindblom, 1990; Kirchner, 1998) or as target undershoot (Bauer, 2008), defined lenition as a means to increase intensity in order to reduce the amount by which the segmental unit interrupts the speech flow (Kingston, 2008). According to this approach, lenition becomes a means to convey the information that lenited units reside inside prosodic constituents, and not at their edges.

Although defining lenition is somewhat difficult, there is a general consensus in the literature regarding which phonetic changes constitute instances of it. They include, amongst others, the spirantization of stops, the opening of fricative consonants into approximants, degemination, debuccalization, and deletion (Kirchner, 1998; Kingston, 2008). Specific examples from Spanish include the weakening of underlying voiced stops /b d g/ into homorganic approximants [β̞ ð̞ ɣ̞] or elision (e.g., Romero Gallego, 1995; Piñeros, 2002), the historical shift from a palatal lateral [ʎ] to [j̞] in most dialects (Sánchez Lobato, 1994; López Gavín, 2015), and the articulation of /s/ as [h] in coda position in several Latin American varieties (Lipski, 1984).

As to variables conditioning lenition, there is ample evidence that speech rate, phonetic context, syllable type and the position of the syllable within the prosodic domain affect the rate of lenition. Higher speech rates normally increase the amount of lenition and elision (Kirchner, 1998; Fosler-Lussier & Morgan, 1999; Dautricourt &

Hume, 2006); however, some speakers are capable of producing unreduced variants at high speech rates (Van Son & Pols, 1990; 1992). In the case of phonetic context, some environments have been traditionally considered strong, such as word-initial, after pause and after nasals, and others weak, such as intervocalic and word-final contexts (Escure, 1977). While strong environments make lenition less likely, the opposite is true for weak contexts.

Several attempts have been made to rank phonetic contexts and syllable types by their relative strength. One often cited example corresponds to the environmental hierarchy proposed by Escure for consonant weakening (1977), according to which consonants in clusters in utterance-final positions are most likely to lenite, followed by word final segments, segments in intervocalic contexts and finally by those in an utterance-initial position[2]. As for prosodic constraints, articulatory evidence has been put forward to show that consonant strengthening (a process that runs in the opposite direction than lenition) is more likely to take place in domain-initial contexts, such as phrase-initial position, as opposed to domain-medial and domain-final contexts, such as inside a phonological word, or at the end of an intonational phrase or utterance, where strengthening is less likely (Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2004; Cho & Keating, 2009).

Lexical frequency has also been shown to have an important effect on lenition. Research has demonstrated that higher frequency morphemes, words and phrases are more likely to be affected by reductive changes, and that these processes will affect high frequency items earlier, but gradually; analogical changes, on the other hand, in which an existing pattern generalizes throughout a corpus, tend to affect low-frequency words first, and are typically categorical (Bybee, 2000; 2002). Examples of phonetic changes occurring first in high frequency domains include cases of spirantization, elision, assimilation and shortening affecting Ethiopian languages (Leslau, 1969), the deletion of word final /t/ and /d/ in contemporary English (Bybee, 2000), the deletion of /d/ in New Mexican Spanish (Bybee, 2002) and the elision of /s/ in favour of [h] in Barranquilla's Spanish (File-Muriel, 2007). The diffusion of sound changes in high-frequency lexical items has been explained as the result of the automation of linguistic productions in casual speech, which ought to favour reduction processes (Bybee, 2002).

2 Segments themselves have also been organized on strong-to-weak hierarchies (see Foley, 1970; Zwicky, 1972).

Another explanation posits that more frequent or already uttered words have less informative weight in discourse, since they are more predictable, and thus speakers utilize less effort to produce them (Fowler & Housum, 1987).

Finally, it has been shown that words are more often affected by reduction processes when they are more predictable or probable (Liberman, 1963; Bell et al., 2003; Gahl, Yao & Johnson, 2012). For instance, the results from a study about shortening processes including the deletion and tapping of /t/ and /d/, and duration shortening in conversational English, showed that knowledge about the likelihood of a given word in context affected how speakers produced words, with higher probability words being more likely to be shortened (Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999). These tendencies have been shown to affect both function and content words (Jurafsky, Bell, Gregory, & Raymond, 2001).

2.2. Cue weighting

Any source of information, in any domain, that allows a perceiver to discriminate between different signals can be said to constitute a *cue* (Toscano & McMurray, 2010). In the case of speech, listeners integrate bundles of cues to discriminate between similar sounds, identifying them and building (more or less abstract) representations of words to accomplish lexical access. Given that sounds are multidimensional, many cues can correlate with a given phonetic category, e.g., place of articulation for fricatives, and every correlate has the potential to act as a cue (Holt & Lotto, 2006; Francis, Kaganovich, & Driscoll-Huber, 2008). Indeed, it is rarely the case that a single acoustic dimension serves to define a category membership unequivocally (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Holt & Lotto, 2006; cf. Stevens & Blumstein, 1981).

The number of potential cues for any given phonetic unit or natural class is large. For example, Lisker reported up to 16 dimensions related to the perception of voicing in English stops (1986). A few examples of well-known cues are oral formants, duration and spectral shape for vowel identification (Delattre, Liberman, Cooper, & Gerstman, 1952; Zahorian & Jagharghi, 1993; Hillenbrand, Getty, Clark, & Wheeler, 1995), voice onset time as a cue to voicing and place of articulation for plosives (Lisker & Abramson, 1964; Cho & Ladefoged, 1999), and spectral peak location, spectral moments, and normalized and relative amplitude for place of articulation in English fricatives (Jongman, Wayland, & Wong, 2000).

Despite the fact that the linguistic information about segmental and suprasegmental levels is primarily encoded in acoustic cues (Chandrasekaran, Sampath, & Wong, 2010), there are several other sources of information that listeners use. For instance, there is ample evidence that speech recognition in noisy environments is aided by the integration of audio and visual cues in face-to-face interaction (Sumby & Pollack, 1954; Erber, 1975; Grant, Walden, & Seitz, 1998; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007), and that there are advantages of audio-visual input for both segmental (Jongman, Wang, & Kim, 2003) and prosodic levels (Swerts & Krahmer, 2008). Also, it has been shown that listeners use cues about phonological context, lexical status and semantic context to recover reduced and highly lenited forms (Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004; Mitterer & Ernestus, 2006), and also use assimilatory cues to anticipate phonetic context (Gow, 2001; Gaskell, 2003).

Both early and recent findings have shown that not every cue is “weighted” in the same way (Mayo & Turk, 2004; Holt & Lotto, 2006; Toscano & McMurray, 2010). There are several reasons why some cues may have more relative importance than others in perception. For example, some cues might provide a better contrast for a given dimension or feature, while others, although present, are not as good at helping to determine category membership (Francis, Kaganovich et al., 2008; Holt & Lotto, 2006). Additionally, some cues may be better encoded in the auditory system, taking advantage of areas of maximum discriminability, receive more attention due to higher variability or are weighted more heavily just for a specific task (Holt & Lotto, 2006). Finally, acquired or learned biases can also account for cue weighting effects (Francis, Kaganovich et al., 2008). In summary, cues are weighted as a function of their reliability, with more reliable cues having more relative weight that unreliable ones (Toscano & McMurray, 2010).

A large body of evidence from developmental and cross-linguistic studies has been put forward demonstrating that the relative weight of the cues for a given phonetic category or phonological contrast is acquired and language-specific (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Polka, 1993; Iverson et al., 2003). Language input has an early effect modulating young infants' ability to discriminate native and non-native categories. This language-specific bias, however, can be modified by linguistic experience and training, both for the segmental (e.g., Logan, Lively, & Pisoni, 1991; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) and suprasegmental level (e.g, Wang, Jongman, & Sereno, 2003; Francis, Ciocca, Ma, & Fenn, 2008).

2.3. Phonological recovery and lexical effects on perception

One of the natural outcomes of lenition is a decrease in the number of available acoustic cues that listeners have for a given phonetic category (Harris & Urua, 2001). In spontaneous and informal speech, the style most frequently encountered by listeners, there is a high prevalence of lenition and elision (e.g., Ingram, 1989; Fosler-Lussier & Morgan, 1999; Johnson, 2004; Torreira & Ernestus, 2011), which compromises the reliability of the acoustic evidence for segments (Ernestus et al., 2002). However, listeners are able to recover lenited and missing segments, provided that secondary acoustic cues and/or additional semantic and syntactical contexts are available (Liberman, 1963; Samuel, 1981a; Samuel, 1987; Samuel, 1996; Mitterer & Ernestus, 2006)[3].

In the case of secondary acoustic cues, studies have shown that listeners are capable of using coarticulatory information in order to recover missing or highly lenited segments. For instance, in an experiment conducted by Yeni-Komshian and Soli (1981), initial fricatives excised from CVCVC sequences were presented to listeners, who were able to recover the following vowel based on formant transitions still present in the fricative consonant. As for syntactic and semantic cues, Kemps et al. (2004) showed that listeners were able to recover absent instances of /l/ from highly reduced samples of the Dutch derivational suffix “*-(e)lijk*” [(ə)lək], often reduced to [ək] or [k] in spontaneous speech, when the reduced forms were presented in a context of several words (see also Ernestus et al., 2002).

3 While this recovery is possible and frequent, reductions do have an adverse effect on word-recognition speed and accuracy (Cutler, 1998; Kemps et al., 2004; Janse, Nooteboom, & Quené, 2007).

Phonemic restoration has also been observed for lenited vowels (Taft & Hambly, 1985), and for segments masked with non-linguistic sounds such as coughs and sinusoid tones (Warren, 1970), even when the transitional sections from neighbouring segments were cueing for a unit different from the target (Warren & Sherman, 1974). Additionally, phonological restoration processes have been shown to interact with lexical status and lexical frequency effects. In the case of the former, more recovery has been found for words as opposed to pseudo-words, although this effect is small (Samuel, 1981a; Samuel, 1996; cf. Samuel, 1987). For lexical frequency, restoration was slightly stronger in high-frequency words than in low-frequency words (Samuel, 1981a; cf. Samuel, 1987 and Samuel, 1981b).

The restoration processes mentioned so far pertain to the recovery of information that is incomplete or entirely missing from the acoustic signal. Another type of perceptual illusion relevant here is lexical effects on the perception of existing acoustic information. Some of the first findings showing the existence of lexical effects on speech perception were observed in word recognition tasks in which faster reaction times were detected in the perception of consonants from words as opposed to nonsense words (Rubin, Turvey, & Van Gelder, 1976). Since then, research on lexical effects on speech perception focussed on the influence of categorical perception, beginning with the well-known study conducted by Ganong (1980), in which he showed that the lexical status of a word –that is, whether it was a word or a nonsense word– had an effect in phonetic categorization; in particular, listeners were biased to interpret an ambiguous acoustic input from phonetic continua as part of words, especially so in and around category boundaries (Ganong, 1980). Similar results showing lexical effects in favour of words on phonetic categorization have been found multiple times, most of them for word-initial segments (Fox, 1984; Connine & Clifton, 1987; Burton & Blumstein, 1995; Pitt, 1995, *inter alia*), but also for stimuli located in word-medial (Connine, 1990) and word-final positions (McQueen, 1991; Pitt & Samuel, 1993).

As a result, phonological recovery and lexical effects on speech processing have become testing grounds for exploration of the architecture of lexical access and of models attempting to account for the transit of information between the acoustic input and lexical representations. For example, findings suggesting that lexical representations mediate the interpretation of acoustic input (present or absent) posit a direct challenge to strong bottom-up models of lexical access such as *Shortlist* (Norris, 1994), in so far as the interpretation of the acoustic signal relies on higher-level sources of information, which allow listeners to recover and reinterpret existing, weakened, masked or missing acoustic input and provide an illusion of perception (Samuel, 1981a; Samuel, 1987; Samuel, 1996; Mitterer & Ernestus, 2006). As to lexical effects, the results have been normally interpreted in support of interactive models of speech perception, such as TRACE (McClelland & Elman, 1986a; 1986b), in which postlexical levels of information can inform lower levels of phonological processing, such as phonetic categorization (Burton & Blumstein, 1995). However, lexical effects have also been said to support autonomous models of lexical access such as RACE (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987), in which separate phonemic and lexical modules process the input simultaneously, and the first one to reach a given threshold provides the basis for a decision.

2.4. Lexical access models

Lexical access is the process by which an acoustic signal is transformed by the listener into a pre-processed input and then mapped into an entry from the mental lexicon of sound images, which is in turn matched to a meaning (Cutler, 1989). In short, lexical access can be thought of as the process responsible for identifying words from acoustic information, which has been traditionally divided into three fundamental stages: *access*, concerned with the forming of a representation of the acoustic input, *selection*, related to the choice of a best matching word-form candidate, and *integration*, the stage at which a recognized unit is related to higher levels of representation (Marslen-Wilson, 1987). The exact way in which this process takes place has been heatedly debated within psycholinguistics since at least the 1960s. As a result, a wide array of lexical access models have been put forward, most of them short-lived, but some surviving after hard theoretical scrutiny, experimental testing and, in most cases, computational implementation. While all lexical access models have the common goal of attempting to provide a system to map the speech signal onto the representations of word forms in the mental lexicon (Marslen-Wilson, 1989), they do show some fundamental differences both in their theoretical assumptions and in the implementation of those assumptions.

Several categories have served to organize lexical access models into families. These groupings are loose, overlap considerably and are not mutually exclusive (Forster, 1989). Perhaps the most widely used classification of lexical access models separates them into *abstractionist*, *episodic* and *hybrid* models, depending on whether they support the existence of intermediate levels of abstract representation of sound units (abstractionist), the storage of multiple episodes or acoustic instances that are later matched directly to lexical representations (episodic), or whether they assume that both types of representation coexist and interact (hybrid). Models can also be characterized according to the directionality of the information flow and influence between the evidence originating from the acoustic input and the information stored as lexical representations. If the information moves only from the acoustic input upwards to higher levels of representation, the model can be characterized as *bottom-up* (which is equivalent to *autonomous*). Instead, if higher levels of abstract representations can affect lower levels of perception the model can be characterized as a *top-down* model, or as an *interactive* model (these two categories, bottom-up and top-down, are not mutually exclusive). Another important way of characterizing lexical access models relates to whether different components or nodes within the model are connected and can interchange information between them or not. Those models in which these links exist are known as *connectionist* models (McQueen, 2005).

In the following sub-sections, the three main lexical access families (abstractionist, episodic and hybrid) will be characterized, and several relevant examples from each will be reviewed briefly.

2.4.1. Abstractionist models

Abstractionist models of lexical access assume that the mental lexicon contains only one abstract representation for each word, and that this representation is made of a string of abstract underlying phonological units (Ernestus, 2014). These are normally assumed to be phoneme-sized, but sometimes other units have been proposed such as syllables, features and articulatory gestures (McQueen, 2005). All models also include a stage of prelexical processing of the signal, in order to extract the relevant phonetic information to inform later stages of processing. Beyond their commonalities, several important differences are observed in the architecture of abstractionist models. To begin with, most models adopt an autonomous approach; that is, there is not top-down influence from information at the lexical level to lower levels of processing. Another area of contention pertains to the nature and number of structures intervening in lexical access processes. For instance, on the one hand, models such as RACE (Cutler & Norris, 1979; Cutler et al., 1987) posit parallel and independent prelexical and lexical processing routes. On the other hand, models like TRACE (McClelland & Elman, 1986a, 1986b) and the *Fuzzy Logical Model of Perception* (FLMP) (Oden & Massaro, 1978; Massaro & Oden, 1980) propose feature, phonemic and lexical processing modules, organized hierarchically. Models like Merge (Norris, McQueen, & Cutler, 2000) are somewhere in between; Merge itself proposes prelexical, lexical and phoneme decision processing nodes, and connections that resemble top-down interaction, but that are strictly bottom-up. Yet another alternative is provided by models like Cohort (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987, 1989; Lahiri & Marslen-Wilson, 1991, 1992) and Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997), in which groups of candidates agreeing to some degree with the acoustic input compete with each other for activation.

Depending on their architecture, perception can take several forms. In broad terms, abstractionist models agree that prelexical decoding modules parse the acoustic input and build preliminary segmental hypotheses, that are then passed along to higher-level modules or are matched directly to underlying abstract representations at the lexical level (after this normalization stage has taken place, the fine acoustic detail is lost). In some simple models like RACE, matching the acoustic input to a lexical representation is binary, but in several models there are degrees of agreement between the prelexical input and candidates, and degrees of activation (e.g., Cohort, FLMP and Shortlist).

Abstractionist models are able to deal with lenition and reduction as long as the lenited acoustic input can be mapped to a unique underlying representation. In most models, for perception, the preprocessed phonetic units are matched to stored lexical representations until one is singled-out depending on the goodness-of-fit and only after surpassing a predefined activation threshold. However, some models like Cohort deal with deviations from the canonical underlying representation by assuming that this representation can be underspecified for some features, given that a word can be accessed even before all the acoustic evidence has been presented (see also Lahiri & Reetz, 2002). A third alternative is provided by models like TRACE, in which lexical and frequency effects, plus phonological recovery, are explained as the result of facilitation from the lexical level to prelexical levels. In production, reduction is simply optional.

Criticism of abstractionist models, particularly extreme ones, comes mainly from evidence showing that listeners cannot ignore talker variability and that knowledge about individual talkers can be stored in long-term memory (McQueen, 2005). As mentioned before, the identity of the talker can influence the performance of participants in a series of experimental tasks (Goldinger, 1998). Also, several frequency and gradedness effects in speech production have considerable experimental support (e.g., Bybee, 2000; Pierrehumbert, 2002).

Review of selected abstractionist models

The *Fuzzy Logical Model of Perception* (FLMP) is an autonomous abstractionist model of speech perception, which attempts to provide a description of the processes involved in the identification of speech sounds based on their phonological features (Oden & Massaro, 1978; Massaro & Oden, 1980). Three separate operations are involved in phoneme identification: *feature evaluation*, *prototype matching* and *pattern classification*. In the first stage of feature evaluation, the degree to which each feature is present in the input signal is determined, and expressed as *fuzzy predicates*[4]. At the second stage of prototype matching, the featural preprocessed results are matched against underlying phonological units and the goodness of each match is determined. Finally, the pattern classification stage determines which phoneme is providing a best match to the input relative to all other phonemes being considered.

TRACE is an interactive abstractionist model of lexical access, developed under the assumption that lexical influences on tasks involving phonemic decisions are a result of

4 Values that can be more or less true instead of only fully true or false.

lexical processes exerting top-down control over prelexical processes (McClelland & Elman, 1986a, 1986b). It proposes three levels of abstract representations: the feature, phonemic and lexical levels. Each one contains a large number of individual *nodes*, which can be thought of as hypothesis detectors, which become activated to a level that is correlated with the strength of the evidence in favour of the hypothesis being tested. Under this model, feature extractors obtain spectral representations of the raw acoustic input, which are passed as features to higher levels dealing with the recognition of underlying phonological units and then to nodes dealing with lexical units. There are bidirectional connections between each level so that information can flow in any direction when required (only when levels share common properties, such as the features that characterize a phoneme and the phoneme itself). Inhibitory connections within each level are modelled as well. Time is represented in this model as a series of networks called *traces*, each one responsible for processing a section of the input and storing the result into an active memory integrated by the concatenation of the networks.

The *Active Direct Access Model*, commonly known as *Cohort*, is an autonomous abstractionist model of lexical access (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987, 1989; Lahiri & Marslen-Wilson, 1991, 1992). Lexical entries in the mental lexicon are posited as being composed of underspecified phonological representations where only unpredictable information is represented. Lexical entries can be activated by appropriate patterns from the acoustic input, but activation is not categorical; instead, activation increases gradually (or decreases) as a function of the goodness of fit of the input to the abstract specification. As an acoustic input begins with a few segments, all the lexical items in the listener's mental lexicon that match the preliminary input become activated simultaneously as candidates, and constitute a “word-initial cohort”. The actual selection of a lexical item is based on the reduction of the number of competitors on the basis of their decreasing activation levels as they begin to differ from the acoustic input. Semantic priming has the effect of facilitating the activation of the target candidate.

RACE, *Shortlist* and *Merge* can be thought of as iterations of the same ideas; in consequence, they will be described as a group here, although in several places in this dissertation they are referenced individually. *RACE* is an autonomous abstractionist model in which access to lexical representations can be achieved by one of two parallel and competing processing routes: a prelexical phonemic level of analysis and a lexical level (Cutler & Norris, 1979; Cutler et al., 1987). Both routes start processing the input at the same time, and the first one to reach a match certainty threshold, or the one being called upon if a task constraints the other, provides the basis for lexical decision. This model predicts that the lexical route will achieve lexical access more often when the input is ambiguous.

Shortlist, also an autonomous RACE model, was created to address evidence showing that lexical feedback to the prelexical processing levels is not necessarily required for lexical access (Norris, 1994; Norris et al., 1997). Under this model, a lexical search is conducted in a word dictionary when an input is received, which derives a short list of best candidates consistent with the bottom-up input[5]. The contents of this list are wired to a lexical network, and the match of the candidates to the incoming input is evaluated as a degree of fit and given a score. Candidate words from the list compete with each other by the means of inhibitory links, whose weight is decided in proportion to the number of segmental units they share (the greater the overlap, the greater their mutual inhibition). Candidates with the lowest bottom-up activation and score are eliminated in an iterative process to make space for other candidates that were not selected initially, and the bottom-up activation score is updated for all candidates until a unique best candidate remains.

Finally, the autonomous model *Merge* (Norris et al., 2000) posits that prelexical processing nodes send information to lexical level decision nodes, which activates compatible lexical candidates. In parallel, the information from the prelexical processing modules is sent to phoneme decision nodes. Finally, information is sent from the lexical level nodes to the phoneme decision nodes, but crucially, not to the prelexical acoustic processing nodes, which warrants a strict bottom-up flow of information. Both at the lexical and phonemic decision levels there are bidirectional inhibitory competition connections between candidates. The phonemic decision nodes receive continuous input from the prelexical processing nodes and the lexical level, and *merge* these two sources of information in order to aid lexical access. All lexical effects on speech perception are thus attributed to the merging of information observed in the phoneme decision units. The absence of inhibitory connections between phoneme nodes at the prelexical level is

5 It is acknowledged that the model lacks a phoneme recogniser module, since phonemic-size units are fed to the dictionary, but they are not derived from raw acoustic input (Norris, 1994).

explained as a mechanism to prevent early categorical decisions that would be difficult to overturn, for example, when the system is exposed to ambiguous input.

2.4.2. Episodic models

The main aspect differentiating episodic models of lexical access from abstractionist models is that the former assume multiple exemplars (also known as episodes, or auditory primitives), each one corresponding to an individual language experience (McQueen, Cutler, & Norris, 2006; Ernestus, 2014). These exemplars, which originate from all experiences in production or perception, are grouped under labels or categories, forming experiential clouds (in strong episodic models, these labels or categories do not have an intervening role in lexical access). Episodic models posit that exemplars contain detailed information about the acoustic and articulatory characteristics of a speech event, including phonetic context. Exemplars can thus be seen as a snapshot of an acoustic event as it develops over time, as a series of spectra or as a spectrogram (Klatt, 1979). However, in some models exemplars are also assumed to contain additional linguistic and non-linguistic information, such as suprasegmental details, characteristics of the speaker's voice, contextual information, attitudes, etc. (Hawkins, 2003). Whichever the assumption about episodes may be, the crucial remaining fact is that episodes are not abstract or normalized; rather, they contain abundant detail from the input signal. Also, no assumption of invariance is required under these models (Klatt, 1979, 1989).

Multiple studies have provided evidence in support of episodic traces in memory. As a whole, they show that listeners pay attention to phonetic detail in the signal, and that listeners can use that source of information to aid perception and lexical access (for an overview, see Hawkins & Smith, 2001). For example, it has been shown that participants are faster in determining if two words in a sequence are the same if they are produced by the same speaker (Cole, Coltheart, & Allard, 1974).

As to the specific relationship between exemplars and lexical access, when an input is being received by a listener during perception, all the exemplars relevant to the input become active. In most models, the degree to which the exemplars will be activated will depend on how well they match the acoustic input (McQueen et al., 2006). Additionally, episodes from highly frequent words and more recent episodes tend to have an advantage over those originating from low frequency words and older exemplars. Retrieval of a lexical item occurs when the activation of a cloud of exemplars is returned to the processing system as an “echo” (Hintzman, 1984, 1986), which contains the result of relating the acoustic input to the stored representations, that is, the label for that group of episodes. Notice that no abstract phonological representation mediates between the episodes and lexical access.

In general, episodic models of lexical access are well equipped to deal with phonetic variation and phonetic change, since they assume that exemplars aggregate naturally into clouds or clusters. Phonetic change can be modelled as a displacement of the exemplars' distribution along the relevant perceptual space, resulting from increasing variability in one direction (bias), decay of older exemplars, and higher availability of more frequent exemplars (cf. Cutler, Eisner, McQueen, & Norris, 2010). Similarly, episodic models are in theory able to deal with lenited and highly reduced word-forms, given that they are stored as any other acoustic experience, and linked to other episodes in proportion to their similarity. However, it seems to be the case that unreduced variants have a preferential status in speech perception, as shown by the fact that processing reduced variants still activates unreduced ones (Ernestus, 2014), and by the fact that reduced variants are not well recognized without additional context (Ernestus et al., 2002; Kemps et al., 2004).

Episodic models of speech perception in general have received criticism on account of their inefficient storage system, in which a huge degree of redundancy exists and heavy demands are imposed on memory (Goldinger, 2007). These issues are addressed through intermediate abstract representations in abstractionist models (McQueen, 2005). Also, evidence from cue weighting is easier to account for if phonological abstract representations are included in the models, because transit from the prelexical level to the abstract representations provides a stage to cue integration and normalization (McQueen, 2005).

Review of selected episodic models

Lexical Access From Spectra (LAFS) is an autonomous episodic model of lexical access (Klatt, 1979, 1989). Lexical representations exist in the form of spectral templates, which receive input from acoustic models resembling a series of spectra. All possible acoustic realizations of a word or word-sequence are precompiled in a decoding network; alternative pronunciations of words and of coarticulated word-boundary segmental units have separate spectral templates. Words are recognized based solely on the match between the acoustic input and the inventory of possible spectral sequences, which is expressed as a distance metric of phonetic differences (e.g., dB differences in frequency bands). Unfamiliar words are processed phonetically by a parallel processor called SCRIBER (for details, Klatt, 1979), and then the new spectrogram is stored in the decoding network. There is no segmental level, feature-level or phonetic analysis mediating between the auditory input and the lexical representations of known words.

Minerva 2 is an interactive model of human memory, in particular, an exemplar model of categorization (Hintzman, 1984, 1986; Goldinger, 1998). This model assumes that all experiences create independent traces in the long-term memory (“episodes”), which store all the details provided by perceptual and contextual sources. In the case of word perception, for every known lexical item, a very large collection of redundant traces are stored. When an input is presented to the system, a retrieval cue, or *probe*, is sent by the working memory to the long-term memory, and all traces relevant to the input are activated in parallel in proportion to their similarity to the input. As a response to this activation, a reply, or *echo,* is returned to the working memory, taking the form of an aggregate of all the activated traces. Given that all relevant traces become activated to some degree, the echo might return information that is no present in the probe, which can be seen of some form of abstraction. However, no intermediate level of abstract representations is posited.

2.4.3. Hybrid models

Hybrid models of lexical access assume that both abstract representations and episodes exist and interact (Ernestus, 2014; McQueen, 2005). They tend to be more recent solutions attempting to address deficiencies from the two contending paradigms, and were built as a response to recent experimental evidence showing that listeners sometimes appear to use exemplars and sometimes abstract representations (McLennan

& Luce, 2005; Mattys & Liss, 2008). Hybrid models share several features, particularly regarding the nature of exemplars, but they do vary in the role they assign to abstract representations. While in some models like Goldinger's Complementary Learning System (CLS) (Goldinger, 2007) and Pierrehumbert's Exemplar Dynamics (ED) approach (Pierrehumbert, 2001, 2002) abstract representations are an integral part of the model, in others like POLYSP (Hawkins & Smith, 2001; Hawkins, 2003) they are a by-product of lexical access and sometimes they are dispensable.

As to their structure, hybrid models present considerable differences. Some models like Goldinger's CLS propose two neuronal networks, one for abstraction and the other one for processing episodes, which interact by the means of intermediate neuronal levels and bidirectional connections. Pierrehumbert's ED model proposes a level of episodic traces associated with labels and phonetic categories, against which acoustic inputs are compared and classified in order to extract a phonological structure to match against lexical representations. POLYSP, finally, gives considerably more importance to multi-modal episodes, that are mapped onto perceptual spaces and organized in order to build linguistic structures and match them to meanings.

In general, since they combine the advantages of both abstractionist and episodic models of lexical access, they are well suited to explain evidence from lexical effects on speech perception, perception of highly lenited forms, learning, phonetic change, etc. Exemplars can account for the phonetic detail being used by listeners and for learning, and abstract representations for speaker-normalization. CLS and POLYSP are also connectionist models, and thus can also account for lexical effects on lower levels of speech processing.

Review of selected hybrid models

Goldinger's Complementary Learning System (CLS) is a connectionist Hybrid model of lexical access (Goldinger, 2007). A “complementary-system” is proposed, in which perceptual behaviour is a result of the interaction of abstract and episodic actions and knowledge. Under this model, two neural networks are characterized: a fast-learning “hippocampal” network and a stable “cortical” network. The former is specialized in processing episodes, and its aim is to memorize specific events fast, as in a working memory. The latter specializes in abstracting and storing statistical patterns and regularities from the input and memory, creating abstract representations for similar stimuli. Between these two neural networks, an “entorhinal cortex” is found, which organizes the information in transit between the other two networks via bidirectional connections that ensure both bottom-up and top-down flows. The acoustic input reaches the cortex neuronal network first, which then sends traces to the hippocampus network. These traces include input from the cortical system, which is in charge of segmenting the input and assigning it meaning, but can also include visual input, emotional data, etc. Consequently, the input that the hippocampal network receives, and from which it is expected to learn specific traces, is already abstract to some extent.

Pierrehumbert's Exemplar Dynamics model of lexical access (and production) stands out from alternative models in that it incorporates mechanisms to deal with synchronic variability and diachronic phonetic change in specific groups of words (Pierrehumbert, 2001, 2002). This model assumes that a large cloud of detailed episodic traces are associated with individual words and phonetic categories, which are abstract in nature (an intermediate level of phonological encoding exists over the exemplar space). The exemplars are organized in a cognitive map in such a way that similar episodes are close to each other. Frequency is implicitly encoded in the structure of the memory system, since frequent categories will be represented by more considerably more tokens. The model also assumes that the parameter space in which the exemplars are represented is “granularized”, which means that the system can only encode phonetic differences larger than the just-noticeable-differences allowed by the perceptual systems. In perception, when an input is encountered, it is classified according to its similarity to existing exemplars. Similarity is computed as its distance from the exemplar in the parameter space.

The *Polysystemic Speech Understanding* model (POLYSP) is a connectionist hybrid model of lexical access (Hawkins & Smith, 2001; Hawkins, 2003). This model gives a central role to phonetic detail, not treating it as noise that has to be normalized, but as a rich source of information. It is also argued that situational context plays a crucial role in understanding the acoustic input. In order to accommodate these claims, it is suggested that phonetic detail is –at least partly– stored in an exemplar memory, which is multi-modal in nature since it includes indexical information about the speaker, emotions, visual cues, context, etc. This memory allows underlying abstract categories to emerge through learning and statistical pattern finding, and is also able to account for the modification of these categories when new information is entered into the system. In order to organize and store both complex linguistic structure and complementary information, the authors use the *Firthian prosodic analysis* paradigm (for details, see Ogden & Local, 1994), which represents phonetic detail in hierarchical structures, including other sources of information. When a listener receives an input, it is mapped onto different parts of the linguistic structure and organized, along with non-linguistic information of all types. Hypotheses about linguistic structures can then be built, and matched to tentative meanings. When a meaning is obtained, the top-down flow of information checks the agreement between the extracted meaning and the input. Abstract representations of the input might be obtained before, at the same time or after lexical access has been attained, or even never, depending on the task at hand.

2.5. Approximant consonants of /b d g/

2.5.1. Defining the term 'approximant'

The first use of the term *approximant* with its current meaning can be traced back to Peter Ladefoged, who in his book *A Phonetic Study of West African Languages* (1968) defined it as referring to “a sound which belongs to the phonetic class of vocoid or central resonant oral […], and simultaneously to the phonological class consonant in that it occurs in the same phonotactic pattern as stops, fricatives and nasals” (p. 25). Before this definition was coined, and even for some time afterwards, what is now called an approximant would have been identified as a fricative in the case of spirant approximants such as [β̞ ð̞ ɣ̞][6].

Ladefoged's definition came about primarily as a means to highlight the phonological contrast in several languages between fricatives and approximants, but also to solve the problem of terminological overlap for two sound classes with differing acoustic characteristics. This overlap stemmed from the fact that the articulatory

6 The term “spirant” distinguishes approximant consonants such as [β̞ ð̞ ɣ̞] from other approximant segments such as semi-vowels, rhotics and laterals (Martínez-Celdrán & Regueira, 2008).

gestures required to produce both manners of articulation are very similar, requiring the active articulator to approach the passive articulator and create some type of constriction without contact. Crucially, however, they differ in that approximants tend to lack the high-frequency turbulent noise which is characteristic of fricatives and of their more narrow constrictions (Ladefoged, 1968, 2003; Maddieson & Disner, 1984; Martínez-Celdrán, 1984, 1991, 2004, 2013; Ashby & Maidment, 2005; Bickford & Floyd, 2006; cf. Romero Gallego, 1995).

Spirant approximants are a manner of articulation that combines several acoustic characteristics of vocoids with the typical phonotactical behaviour of consonants. Given that in their production the articulators only approach, the air flow is able to produce resonances more or less as it would for vowels, and some oral formants –normally only F1 and F2– are present in the signals (Ladefoged, 2003). Approximant consonants are the least constricted amongst consonants, but they do display higher degrees of constriction than semi-vowels (Martínez-Celdrán, 2004). They are also known for being relatively short, displaying brief formant transitions and an intensity decrease with respect to neighbouring segments (Laver, 1994; Colantoni & Marinescu, 2010). Phonotactically, approximants tend to be located at syllable onsets and codas, which is why they are considered consonants (Martínez Celdrán, 2013).

It has been proposed that approximant consonants form a supra-category that includes semi-vowels, spirant central approximants, rhotic central approximants and laterals (Martínez-Celdrán, 2004). Furthermore, three sub-categories have been proposed for spirant central approximants: vocalic, open and closed approximants (Martínez-Celdrán, 2004, 2013; Martínez-Celdrán & Regueira, 2008). These three categories distinguish sub-types of spirant approximants in terms of their degree of constriction, with vocalic approximants as the less constricted units, and closed approximants as the most constricted units (e.g., Martínez-Celdrán, 2004). In this dissertation, the term *approximant* will be used to refer to spirant central approximants, and in particular to refer to the approximant variants of Spanish /b d g/, which will be transcribed as [β̞ ð̞ ɣ̞][7].

7 Several other spirant approximant variants have been reported for Chilean Spanish. For example, [ʝ̞] from /ʝ/ (Sadowsky, 2015) and [ɹ] from /r/ (Sadowsky & Salamanca, 2011).

2.5.2. Spanish /b d g/ and their approximant variants

The traditional account for Spanish /b d g/ states that they constitute a natural class in the phonological system in opposition to the voiceless series /p t k/; plosive variants [b d g] –bilabial, postdental and velar, respectively– are found after pause, nasals and after /l/ in the case of /d/; elsewhere, in complementary distribution, these consonants would be articulated as the approximants [β̞ ð̞ ɣ̞] (Hualde, 2005). Still in the context of the traditional account, the process by which voiced stops lenite to approximants is usually termed *spirantization* (Cole, Hualde, & Iskarous, 1999).

In very broad terms, this general description can be said to hold for the majority of Spanish dialects, but deviations from this account have been reported with increasing frequency in recent years. For example, some dialects, e.g., varieties in Central America and Colombia, seem to have undergone fortition processes and produce voiced stops in contexts where normally approximants would have been expected, i.e., all but intervocalic position (Amastae, 1989; Hualde, 2005; Carrasco, Hualde, & Simonet, 2012; Harper, 2014). Examples in the opposite direction are Chilean and Miami Spanish, in which the degree of lenition and elision are particularly high (e.g., Pérez, 2007; Hammond, 1967). Of course, there is important variation within dialects as well, with all of them displaying a spectrum or continuum between very constricted variants to fully vocalized or elided variants, in different proportions, which contradicts a binary account of oral stops opposed to approximants, as the complementary distribution account would suggest (Almeida & Pérez Vidal, 1991; Cole et al., 1999; Martínez-Celdrán & Regueira, 2008; Hualde, Simonet, Shosted, & Nadeu, 2010; Hualde, Shosted, & Scarpace, 2011; Carrasco et al., 2012; Simonet, Hualde, & Nadeu, 2012). Moreover, while /b d g/ form part of a natural class, they do display clear differences with /g/ normally being the less lenited category and /b/ or /d/ being the most lenited, depending on the specific dialect: /b/ in the case of Madrid and Costa Rican Spanish (Carrasco et al., 2012), /d/ in the case of the dialects from Argentina and Chile (Colantoni & Marinescu, 2010; Pérez, 2007), and both /b/ and /d/ with similar rates of lenition for Colombian and Mexican variants (Harper, 2014).

A number of variables have been shown to have an effect on variation in /b d g/. One of the most thoroughly studied is the role of phonetic context and stress. In the case of phonetic context, it is clear that word-medial instances –and particularly so for intervocalic contexts– are more lenited than word-initial instances, even for Spanish dialects affected by fortition processes (Martínez Celdrán, 1984; Hualde et al., 2010; Eddington, 2011; Carrasco et al., 2012; Simonet et al., 2012; Harper, 2014). On the other hand, more constricted realizations of approximants and stops are found after pauses, nasals, fricatives, and laterals (Hualde et al., 2010; Eddington, 2011; Hualde, Shosted et al., 2011; Simonet et al., 2012; Harper, 2014). Madrid, Costa Rican, Colombian and Mexican Spanish variants display more lenition when the consonants are located before /a/ (Carrasco et al., 2012; Harper, 2014), while Argentinian Spanish shows more lenition before back vowels for /d/ (Colantoni & Marinescu, 2010) and Castillian Spanish more lenited variants before front vowels for /g/ (Cole et al., 1999). In the case of lexical stress, more lenited variants are always found after a stressed vowel or syllable, as opposed to in the onset of a stressed syllable (Cole et al., 1999; Ortega-Llebaria, 2003; Hualde, 2005; Eddington, 2011; Carrasco et al., 2012; Harper, 2014), particularly for /d/ (Colantoni & Marinescu, 2010).

Lexical frequency has also been linked to the degree of lenition for Spanish voiced plosives, although only for /d/. In short, more lenition for /d/ has been found in lexical items or lemmas with higher lexical frequencies (Bybee, 2002, 2003; Eddington, 2011; Brown, 2013), and in the highly frequent past participle suffix “-ado” (Bybee, 2003).

Several studies have reported results describing the acoustic characteristics of spirant approximants of Spanish /b d g/. However, it is not always clear how the segmentation was carried out and how the acoustic measurements were extracted, and very few of these studies report the use of normalization protocols to control for confounding variables such as speech rate or physiological differences between the speakers' vocal tracts. Consequently, these results should be interpreted with caution. In terms of duration, approximant consonants have been found to have a short average duration, ranging between 30 and 45 ms according to some accounts (Almeida & Pérez Vidal, 1991) and around 40 to 60 ms on others (Martínez Celdrán, 1984, 2013). As a general trend, the more lenited the unit, the shorter its duration (Martínez Celdrán, 1984). Approximant variants from /d/ tend to display the shortest durations (Colantoni & Marinescu, 2010). As for intensity, several recent studies have used relative intensity measures as correlates of degree of constriction (Cole et al., 1999; Ortega-Llebaria, 2003; Colantoni & Marinescu, 2010; Hualde et al., 2010; Eddington, 2011; Hualde, Shosted et al., 2011; Hualde, Simonet, & Nadeu, 2011; Carrasco et al., 2012; Simonet et al., 2012), or to normalize intensity (Martínez Celdrán, 1984, 2013), finding that the intensity of less constricted variants is more similar to that of the surrounding segments (i.e., more lenited approximants are more "intense") and that they have lower velocity of intensity change. That is, the change in intensity with respect to neighbouring segments is less abrupt when compared to more constricted variants. In the case of oral formants, average F1 values of 405 Hz were found for [β̞], 383 Hz for [ð̞] and 446 Hz for [ɣ̞] (Almeida & Pérez Vidal, 1991). Average F2 values of 1080 Hz were found for the bilabial approximant, 1360 Hz for the postdental and 1368 Hz for the velar variant (Almeida & Pérez Vidal, 1991).

Only a few recent studies have approached variation of Spanish /b d g/ from an articulatory perspective. An electropalatography study aiming to characterize the degree of constriction for /d/ variants showed that more occlusion was found for /d/ after /l/ and /n/, while no evidence of occlusion was found for /d/ after /r/, /s/ or vowels (Hualde, Shosted et al., 2011). In a similar pilot study using electropalatography, Hualde et al. (2010) showed that /d/ always surfaced as [ð̞] after the vowel /a/ and /r/, and that the plosive variant was found after nasals and laterals; no difference was observed for the degree of constriction for /d/ after /n/ or /s/.

Surprisingly little is known about how listeners perceive phonetic variants of Spanish /b d g/, besides studies exploring the perception of voice-onset timing in stops (e.g., Abramson & Lisker, 1972; Williams, 1977; Zampini, 1998). To the best of our knowledge, the only study that has tested perception of Spanish approximant variants was conducted by Harper (2014), who set out to study the role of perception as a mechanism of sound change in the spread of spirantization. The author had native speakers from low and high spirantization dialects react to the contrast between plosive and approximant variants of /b d g/ in several identification, discrimination and word repetition tasks. The results showed that discriminability between plosives and approximants was higher in stressed syllables, that place of articulation was better discriminated in stop realizations, and that there was a preference for stop-like articulations in word-initial contexts in a repetition task with nonsense words. On the whole, it was concluded that the (mis)perception of spirantization was not a mechanism for sound change in Spanish.

2.5.3. Chilean Spanish /b d g/

The distinction between fricatives and approximants is relatively new in the specialized literature for Spanish (the first mention for Spanish corresponds to Martínez Celdrán, 1984; and in Chilean Spanish probably Cepeda, 1991). Unless indicated otherwise, I will assume here that authors reporting fricative allophones of /b d g/ are referring to approximant variants. I base this decision on three arguments:

(a) Most authors do not mention approximant variants in their descriptions. Given that the terminological distinction between fricatives and approximants was only truly introduced into the Spanish research community by the late 1990s or even later, it is safe to assume that “fricative” can be replaced by “approximant” when it pertains to variants of /b d g/.

(b) True voiced fricatives of /b d g/ are an oddity in contemporary Chilean Spanish (see Chapter 4).

(c) Voiced fricatives are aerodynamically challenging (Ohala, 1983), and thus it is difficult to argue in favour of a lenition process in which voiced stops go through a voiced fricative stage on their way to elision (although, see Piñeros, 2002).

Phonetic variants of /b/

Several phonetic variants have been reported for Chilean Spanish /b/. The plosive variant [b] is normally articulated as such after pause and homorganic nasal consonants (Lenz, 1940b; Oroz, 1966; Salas, 1996-1997; Cepeda, 2001; Borland Delorme, 2004; Sadowsky, 2010). Elsewhere, other allophones of /b/ predominate. Fricatives have been reported as well (Lenz, 1940b; Cepeda, 1991; Borland Delorme, 2004; Sadowsky, 2010) and “relaxed fricatives” (Oroz, 1966; Cepeda, 1991). As mentioned above, these variants are particularly frequent after unstressed vowel and intervocalically (Cepeda, 1991). Approximant variants are favoured in intervocalic contexts, in syllable-initial position with other consonants and after vowels (Silva-Fuenzalida, 1952-1953; Salas, 1996-1997; Borland Delorme, 2004). Finally, “very relaxed” variants are mentioned in older references (Lenz, 1940b; Oroz, 1966).

As described in detail by Sadowsky (2010), in general, labiodental variants of /b/ were considered non-existent in Chilean Spanish, until around 1990. For instance, Oroz (1966) explicitly denied the existence of labiodental variants of /b/. However, labiodental variants have been reported several times since then. Cepeda (1991), when describing the phonetic system of Valdivia's Spanish –a city located in southern Chile–, mentions a voiced labial-dental fricative [v] and a voiced dorsum-velar fricative lenis [ʋ̯]. Borland Delorme (2004) also mentions [v], without specifying whether there is a phonetic context favouring that variant. Sadowsky (2010) describes approximant and fricative labiodental variants for /b/, although there is no attempt to differentiate between these two manners of articulation in his study; the same variants ([v], [ʋ]) are reported by Sadowsky and Salamanca (2011). Labiodental variants have been shown to be the most frequent variant of /b/ (Sadowsky, 2010; Vergara Fernández, 2011, 2013; Vergara & Pérez, 2013), present in a wide array of phonetic contexts, both in syllable onset and coda, after liquids and unrounded vowels, although word initially and after nasal bilabial variants are more common (Sadowsky, 2010; Vergara Fernández, 2011). Finally, several reports have been able to rule out the hypothesis that the orthographic distinction between <b> and <v>, learned through literacy, conditions the place of articulation for instances of /b/ (Sadowsky, 2010; Vergara Fernández, 2011; 2013; Vergara & Pérez, 2013).

Elided variants of /b/ are also frequently reported (Véliz, Araya, & Rodríguez, 1977; Cepeda, 1991, 1994; Cepeda & Poblete, 1993; Cid & Céspedes, 2008). Elided variants are more frequent in males (Cepeda, 1991), when /b/ is located in a functional morpheme (Cepeda & Poblete, 1993), in intervocalic contexts (Cid & Céspedes, 2008), after stressed vowels, after approximant consonants and in the syllabic coda (Cepeda, 1991).

Phonetic variants of /d/

Plosive realizations are mostly found after pause, /n/ and /l/ (Oroz, 1966). Elsewhere, fricative, approximant and elided variants are normally articulated. As was the case for /b/, it remains unclear whether authors describing fricative variants were not actually referring to approximant variants, and thus references to fricatives should be considered cautiously. Several fricative variants have been reported, among them, voiced alveolar fricatives (Oroz, 1966) and voiced dental fricatives (Cepeda, 1991). Approximant variants have been reported more recently by Contreras (1993), Salas (1996-1997) and Borland Delorme (2004). These variants are facilitated by intervocalic contexts and are frequent after vowels (Silva-Fuenzalida, 1952-1953), or any other context that does not facilitate [d] (Salas, 1996-1997). The most frequent variant is [ð̞] (Cepeda, 1991). Several authors sometimes refer to particularly "relaxed" or "loose" articulations of /d/ (Oroz, 1966; Rabanales, 1992, 2000; Valdivieso, 1993). These variants probably correspond to approximant articulations, perhaps vocalic approximants. Only recently there have been reports of interdental variants of /d/, both fricative and approximant (Sadowsky & Salamanca, 2011; Sadowsky, 2015).

Elision of /d/ in Chilean Spanish is well reported in the literature. It was first reported by Bello (1940: 52, 53), who mentions that the elision of /d/ should be avoided in nouns and adjectives ending in "*-do*" and "*-dos*", and is further mentioned by e.g., Lenz (1940a), Véliz et al. (1977), Wigdorsky (1978), Cepeda (1991, 1994), Rabanales (1992, 2000), Cepeda and Poblete (1993), Contreras (1993), Cid and Céspedes (2008) and Sadowsky (2015). The phoneme is reported to be most frequently elided in intervocalic contexts and /d/ at syllable coda when followed by consonant (Wigdorsky, 1978; Rabanales, 1992, 2000; Contreras, 1993). Elision in the syllable coda, after a stressed vowel (Cepeda, 1991), in functional suffixes (Cepeda & Poblete, 1993; Quilis, 1999) and in word-final position (Rabanales, 2000; Cid & Céspedes, 2008) is also mentioned.

Phonetic variants of /g/

According to the literature, plosive variants of /g/ are normally found after pause and nasal consonants (Silva-Fuenzalida, 1952-1953; Oroz, 1966; Cepeda, 2001), while fricative and approximant variants are common elsewhere (Borland Delorme, 2004; Salas, 1996-1997), although also sometimes after a pause (Silva-Fuenzalida, 1952-1953). Some fricative variants mentioned –probably approximants– are a voiced dorsum-velar fricative and a lenis version of the same segment (Cepeda, 1991). As for approximants, a voiced velar approximant [ɣ̞] is reported (Sadowsky & Salamanca, 2011), along with a voiced labial-velar approximant ([w]) has been reported as an allophone of /g/ (Cepeda, 1991). One aspect that differentiates the variants of /g/ from the other phonemes in the series is that all the velar variants are affected by a coarticulatory palatalization process when preceding the front vowels /e/ and /i/ (Oroz, 1966; Borland Delorme, 2004). This phoneme tends to display lower degrees of elision than /b/ and /d/ (Lenz, 1940a). However, elided variants are typical before /u/, when forming a diphthong with [w] (Lenz, 1940b), between vowels, before a pause (Cepeda, 1994) and before approximants (Cepeda, 1991).

Summary

As is the case for other varieties of Spanish, Chilean Spanish /b d g/ have been traditionally reported to be realized as plosives after a pause, homorganic nasals and as /d/ after /l/. Elsewhere, fricative, approximant and elided variants are found, with approximants as the most frequent (it remains unclear whether the authors that mention fricative variants are referring to approximants, especially in early studies). Recent studies show that, despite clear trends, there is high variability in the realization of /b d g/ in all contexts.

It is fair to characterize the variants of /b/ as predominantly labiodental and then bilabial, those of /d/ as postdental, and the variants of /g/ as velars or as palatalized velars when preceding front vowels. Elision is reported consistently in several studies, for the whole series. Elision is particularly high for /d/, followed by /b/ and then /g/ (Pérez, 2007), with elision particularly common intervocalically, after stressed vowels, after approximant consonants and in the syllabic coda and word-final positions. Elision was also found to be particularly high in some functional suffixes in the case of /b/ and /d/. There is no mention in the literature for Chilean Spanish of a continuum of variants from approximants to elision, although it is clear that this is the case given the reported variation.

2.6. The current thesis

As stated at the beginning of this chapter, this dissertation can be summarized as having three main objectives: determining the scope of lenition in Chilean Spanish /b d g/, testing how listeners handle lenited variants in perception, and evaluating how the results from these two studies agree with traditional and current models of lexical access and speech perception. As to the first objective, the reviewed literature for the approximant variants of /b d g/ showed that, first, lenited and elided variants are an integral part of the allophony for the voiced plosive series of Spanish, although it remains contested whether plosive variants lenite to approximants or the opposite process takes place. It is also clear that different dialects of Spanish can display very different patterns of variation in the fortis-lenis continuum, with some variants showing an allophonic variation consistent with a fortis account (e.g., Costa Rican Spanish and some dialects of Colombian Spanish), and others displaying high degrees of lenition and elision (e.g., Miami and Chilean Spanish). Additionally, the degree of lenition in Spanish /b d g/ is facilitated in some phonetic contexts and in higher frequency words. Generally speaking, approximant consonants are short, and they always have lower intensity than their neighbouring segments.

Except for the fact that lenition and elision are found in predictable phonetic contexts (e.g., intervocalically), little is known about Chilean Spanish approximant consonants. Moreover, most studies were conducted a while ago, and may not reflect current usage. Crucially, no study provides acoustic data that perception experiments with synthetic stimuli can use as reference values, or against which results can be interpreted. Chapter 3 and Chapter 4 present the results of the first large-scale production study of Chilean approximant variants of /b d g/. Chapter 3 summarizes the aims and methods used in the production study whilst Chapter 4 gives a detailed account of the findings, focussing not only on raw acoustic data, but also on the role of indexical variables such as phonetic context, word status, experimental task and word frequency in the production of these variants. The information collected in the production study and a small follow-up perception study in which the ability of listeners to discriminate between bilabial and labiodental approximant variants of /b/ is investigated (Chapter 5) will provide a detailed account of the full scope of lenition and elision in Chilean Spanish, including an acoustic characterization of the consonants. These results provide the basis for subsequent experiments investigating the perception of lenited and elided variants of /b d g/, described later in Chapter 6 and Chapter 7.

Chapter 3
Production of /b d g/: Introduction and methods

3.1. Introduction

The general aim of this chapter and Chapter 4 is to determine the scope of lenition in Chilean Spanish /b d g/. In order to achieve this, a full acoustic characterization of the approximant variants of /b d g/ will be provided, along with a statistical exploration of some variables conditioning this variation. While these objectives might appear straightforward at first, the continuous nature of spirant approximant consonants imposes a series of methodological challenges, which will be addressed in this chapter.

3.1.1. Summary of state of the art

There are several reasons that explain the relative abundance of studies addressing the allophonic variability of /b d g/ in Spanish. Firstly, the spirantization versus stop-formation debate is a traditional and very well known issue within Spanish phonology. This issue is still relevant today because it highlights the interface between phonetics and phonology, and at the same time challenges the theoretical assumptions of many phonological paradigms. Secondly, this debate has direct bearing on the assumptions about how phonetic and phonological change occurred in the evolution from Latin to early Romance languages and then onwards until modern Spanish /b d g/ (e.g., Lozano, 1978; Lapesa, 1981). Thirdly, as reviewed in “2.5.1. Defining the term 'approximant'”, the articulatory and acoustic characteristics of [β̞ ð̞ ɣ̞] posed problems for some definitions of certain manners of articulation, prompting discussions on the appropriate place of spirant approximants in a literature dominated by accounts of plosives and fricatives. Finally, the variation of /b d g/ has important implications for the Spanish phonological system, particularly regarding the opposition between /b d g/ and other natural classes such as /p t k/ (Almeida & Pérez Vidal, 1991; Hualde, Simonet & Nadeu, 2011; Parrell, 2011).

It is well known that /b d g/ are realized as plosives in strong phonetic contexts (e.g., after nasals), and that other variants including spirant approximants and elided variants appear in contexts facilitating lenition (e.g., Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012). It is also well established that this variation is dialect dependent, in that some variants of Spanish currently show evidence of fortition, while others run in the opposite direction (e.g., Carrasco, Hualde & Simonet, 2012; Pérez, 2007). There is considerable variation within dialects too, and it has become clear that variable rules with binary or categorical outputs fail to convey the true extent of the variation in Spanish /b d g/. When /b d g/ are compared, some important differences are found. Across dialects, /b/ and /d/ variants are normally more lenited than /g/ (e.g., Pérez, 2007; Colantoni & Marinescu, 2010; Harper, 2014), but it is less clear which of /b/ or /d/ displays more lenited variants.

Several variables condition the variation of /b d g/. To summarize, prominent phonetic contexts favour more *fortis* variants (e.g., Cole, Hualde & Iskarous, 1999; Ortega-Llebaria, 2003; Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012), word-frequency is positively correlated with degree of lenition (Bybee, 2002, 2003; Eddington, 2011; Brown, 2013), and amount of attention to speech or degree of formality in an elicitation procedure are negatively correlated with degree of lenition (e.g., Carrasco, Hualde & Simonet, 2012). However, not much is known about the acoustic and articulatory characteristics of the approximant variants of /b d g/. Still less is known about how they are perceived by listeners. In the acoustic domain, data has shown that approximant variants are relatively short (e.g., Martínez Celdrán, 2013), that degree of lenition is negatively correlated to duration (Martínez Celdrán, 1984), and that the intensity differential between the approximant consonant and its neighbours decreases as degree of lenition increases (e.g, Hualde, Shosted & Scarpace, 2011; Simonet, Hualde & Nadeu, 2012). A couple of studies also offer non-normalized reference formant values (e.g., Almeida & Pérez Vidal, 1991). In the case of the articulatory domain, some evidence has been put forward in agreement with the assumptions that more stop-like realizations of /d/ are found after laterals and nasals (Hualde, Shosted & Scarpace, 2011), and that more lenited variants surface after vowels and rhotics (Hualde, Simonet, Shosted & Nadeu, 2010).

Existing descriptions of Chilean Spanish /b d g/ do not deviate much from this general account. The only differences of importance are the fact that elided variants seem to be particularly frequent (Cepeda, 1991, 1994; Cepeda & Poblete, 1993), and that labiodental realizations have been reported for /b/ (Sadowsky, 2010). In the case of Chilean Spanish, there is no acoustic data available for the approximant variants of /b d g/. No articulatory or perceptual studies exist either.

3.1.2. Methodological standards

The main methodological challenges facing the acoustic study of spirant approximant consonants originate from two separate domains: segmentation and normalization. In the case of segmentation, spirant approximant variants pose particular difficulties because they display continuous formant transitions with their surrounding segments (Martínez-Celdrán & Regueira, 2008), and as there is no obvious abrupt transition between these segments and their neighbours, this makes defining a boundary an arbitrary decision, at least to some degree.

Two approaches exist as a response to the segmentation problem. The first one consists in acknowledging and accepting the subjective nature of manual segmentation, and performing manual segmentation (e.g., Kingston, 2008). A second approach, which only works for some research questions, consists of completely bypassing the segmentation problem by using relative intensity measurements, obtained automatically via scripts, to investigate topics related to constriction degree and elision (e.g., Eddington, 2011; Hualde, Shosted & Scarpace, 2011; Hualde, Simonet & Nadeu, 2011; Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012). Besides the obvious problems surrounding the first approach, the alternative has opposite advantages and disadvantages. Manual segmentation has the advantages of being conducive to isolating a linguistic variable (i.e., the segment), and thus allowing extraction of acoustic information, including duration; on the other hand, manual segmentation is a highly subjective process and highly time-consuming. Relative intensity measurements have the advantages that they can be obtained automatically (once a quick preliminary manual segmentation has been completed), and consequently the process takes considerably less time, and is fully objective; on the other hand, no linguistic unit becomes isolated in the process, and arguably not much can be said about other acoustic characteristics besides intensity. In this dissertation, manual segmentation was chosen because no forced aligner has been trained in Chilean Spanish, and also because duration measurements are to be collected, and thus semi-automated methods cannot be employed exclusively.

A similar picture is observed regarding normalization. Acoustic measurements from vocoids are known to be affected by a number of confounding variables, and thus normalization is required to remove these sources of undesired variation. In the case of duration, it is directly conditioned by speech rate, which is known to be sensitive to variables such as the participant's dialect and age (Jacewicz, Fox, O'Neill & Salmons, 2009), gender (Jacewicz, Fox & Wei, 2010), the elicitation procedure (Barik, 1977), and to even vary within speakers (Jacewicz, Fox & Wei, 2010). In the case of intensity, undesired sources of variation include fluctuations in the distance between the participant's mouth and the microphone (Titze & Winholtz, 1993), and the overall intensity with which different speakers produce speech (Hodge, Colton & Kelley, 2001). As to oral formants, physiological differences between speakers affect measurements (Peterson & Barney, 1952).

When it comes to solutions to undesired sources of variation, different studies reporting acoustic data for spirant approximant variants have taken different approaches. Some studies aim to provide absolute reference values, and consequently no normalization procedures are applied (e.g., Martínez Celdrán, 1984). Other studies apply normalization procedures in order to control some of the confounding variables mentioned above (e.g., Martínez Celdrán, 2013). A final group of studies, uses normalization for purposes different from providing acoustic values, in particular, as correlates of degree of constriction (e.g., Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012).

This chapter addresses the methodological problems surrounding segmentation and normalization, and presents an explicit and consistent procedure that is applied in a study of Chilean approximant consonants.

3.2. Aims and objectives

This study aims to address several research questions. First of all, given the lack of current acoustic data for Chilean Spanish [β̞ ð̞ ɣ̞], it aims to investigate whether highly lenited and elided variants predominate, as might be expected based on previous literature. Besides its theoretical relevance for the lenition versus fortition debate (Harris, 1969; Lozano, 1978; Danesi, 1982; Mascaró, 1984; Baković, 1994; Piñeros, 2002; Barlow, 2003), and the language change literature (e.g., Lozano, 1978; Lapesa, 1981), this has direct implications for perception, given that this variation is expected to modulate listeners' expectations regarding what constitutes expectable instances of /b d g/. For example, if lenition and elision are widespread, listeners might be particularly sensitive to weak acoustic evidence cueing for the presence of an approximant consonant, in order to aid in the discrimination of minimal pairs such as *presidente* [pɾe.si.ˈð̞en.te] (“president”) versus *presiente* [pɾe.ˈsjen.te] (“he foresees”). Another possibility is that listeners might not be sensitive to small acoustic differences at all, but instead highly dependent on alternative sources of evidence, such as semantic or syntactic contexts.

Another area that requires confirmation is the role of independent variables such as phonetic context, word status, elicitation procedure and word frequency in the variation of /b d g/. Determining their influence is not only interesting in that it provides updated evidence to help explain the variability in /b d g/, but in that it has direct bearing on the design of perception experiments. Phonetic context, word status and word frequency are particularly important, given that any stimuli built to explore the perception of [β̞ ð̞ ɣ̞] will necessarily manipulate these three variables, and all of them have been shown to have an effect on perception (see sections “2.1. Lenition”, “2.2. Cue weighting” and “2.3. Phonological recovery and lexical effects on perception”).

Finally, a number of authors have pointed out the gradient nature of the variation of Spanish /b d g/ from full approximants to elided variants (Almeida & Pérez Vidal, 1991; Cole, Hualde & Iskarous, 1999; Hualde, Simonet, Shosted & Nadeu, 2010; Hualde, Shosted & Scarpace, 2011; Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012), but evidence has yet to be presented confirming this for Chilean Spanish. If continua are not substantiated by acoustic evidence, then using them in perception tasks might diminish the ecological validity of any experiment. However, if continua are found, a number of additional research questions arise, e.g., whether listeners parse each continuum categorically or not.

Primary aims

(a) Determine the scope of lenition in Chilean Spanish /b d g/.
(b) Confirm the role of phonetic context, word status, elicitation procedure and word frequency in the variation of Chilean Spanish /b d g/.
(c) Determine whether an acoustic continuum from approximant consonant to elision exists in Chilean Spanish /b d g/.
(d) Provide reliable acoustic reference values of Chilean Spanish approximant consonants to inform the preparation of stimuli for perception experiments and to aid the interpretation of their results.

Secondary aims

(a) Determine whether the classification into vocalic, open and closed approximants is backed-up acoustically in Chilean Spanish /b d g/.
(b) Compare the variation of Chilean Spanish /b d g/ to that of other dialects.
(c) Contribute to setting methodological standards for the description of the acoustic variation of spirant approximant consonants in general, and Spanish [β̞ ð̞ ɣ̞] in particular.

3.3. Data collection

3.3.1. Participants

Ten participants (5 male, 5 female; mean age 21.1) took part in the elicitation tasks. All participants were adult native speakers of Chilean Spanish, studying undergraduate or graduate programmes, and residents of Santiago (Chile). Participants had not lived outside Chile for any significant amount of time. No speech, hearing or cognitive impairment was reported by any of the participants. Participants were paid for their participation.

3.3.2. Elicitation procedures

Speech samples were obtained by the means of three elicitation tasks: word-lists, short texts and semi-guided conversations. These three elicitation procedures were used under the assumption that they provide speech samples subject to varying degrees of perceived formality and attention to speech (Labov, 1972). Word-lists consisted of 84 words and 51 nonsense words, in which approximant consonants were embedded in five intervocalic contexts and three stress patterns (see Table 3.1 for details). The words used were extracted from the LIFCACH word frequency list (Sadowsky & Martínez Gamboa, 2004). High frequency items were prioritized whenever possible. Tokens were embedded in a low predictability carrier phrase – "Diga ____ cada vez" (Martínez Celdrán, 1984) – and grouped into blocks of 20 items, after randomization. Participants were instructed to familiarise themselves with each block before reading it aloud twice, and to read them as naturally as possible. In the event of a mistake, participants were asked to repeat that token at the end of the recording session.

In the second task, participants read aloud two short texts resembling news articles (see Table 3.2). Care was taken to balance the number of words containing /b/, /d/ and /g/ instances, in a variety of phonetic contexts. Participants were again asked to familiarize themselves with the texts before reading them aloud for recording; in case of mistakes, the participants were asked to repeat the relevant section at the end of the recording session.

The last elicitation procedure was a semi-guided conversation between a phonetically trained researcher and the participant. The task began with a conversation about hobbies and then it progressed to other subjects of interest to the participant until at least 10 continuous minutes of speech had been produced and recorded.

Table 3.1. List of words and nonsense words for the first elicitation procedure. Five intervocalic contexts and three stress positions relative to the target approximant consonant were included for each consonant. Target consonants have been underlined, translations to English are provided for words in brackets, and nonsense words are in italics.

Phoneme	Vocalic context	Following stress	Stress on syllable	Preceding stress
/b/	/i/	Valdivia[8] (a Chilean city) Bolivia (Bolivia) *tíbi*	recibir (to receive) vivir (to live) *tibí*	posibilidad (possibility) actividad (activity) *tibitá*
	/e/	jueves (Thursday) breve (brief, short) *tébe*	deber (duty) evento (event) *tebé*	revelar (to reveal) eventual (eventual) *tebetá*
	/a/	sábado (Saturday) Hormazával (a surname) *tába*	trabajo (job) acabar (to finish) *tabá*	trabajar (to work) avanzar (to advance) *tabatá*
	/o/	robo (robery) lobo (wolf) *tóbo*	robot (robot) soborno (bribe) *tobó*	provocar (to provoke) corroborar (to corroborate) *tobotó*
	/u/	úvula (uvula) búbu (a brand of sweets) *túbu*	suburbio (suburb) suburbios (suburbs) *tubú*	suburbano (suburban) tubular (tubular) *tubutú*
/d/	/i/	jurídico (legal, juridical) fatídico (fateful) *tídi*	decidir (to decide) coincidir (to coincide) *tidí*	tridimensional (three-dimensional) ridiculizar (to ridicule) *tiditá*
	/e/	sede (headquarters) Mercedes (a name) *téde*	antecedente (precedent) suceder (to happen) *tedé*	alrededor (around) federación (federation) *tedeté*
	/a/	cada (each) nada (nothing) *táda*	ciudadano (citizen) Canadá (Canada) *tadá*	década (decade) adaptar (to adapt) *tadatá*
	/o/	todo (all) modo (mode) *pódo*	Almodóvar (a surname) Rodolfo (a name) *todó*	método (method) cómodo (comfortable) *todotó*
	/u/	*lúdus* *dúdu* *túdu*	pudú (a type of dear) vudú (vudu) *tudú*	fraudulento (fraudulent) mapudungún (mapuche language) *tudutú*
/g/	/i/	amigui[9] (friend) amiguis (friends) *tígui*	liguilla (league) amiguito (small friend) *tiguí*	Diguillín (a brand) liguillero (related to leagues) *tiguitá*

8 In Spanish, both characters “b” and “v” map to the underlying category /b/, which can be instantiated by bilabial or labiodental allophones. Bilabial and labiodental variants occur in free variation in Chilean Spanish (Sadowsky, 2010), and no systematic correlation between grapheme and place of articulation has been found (e.g., Vergara Fernández, 2013).

9 In Spanish spelling, the vowel “u” after “g” is not pronounced. Its purpose is to make explicit that “g” will be pronounced as [g] instead of [x] when “g” precedes “e” or “i”.

Phoneme	Vocalic context	Following stress	Stress on syllable	Preceding stress
	/e/	despliegue (display) despegue (take-off) *tégue*	ceguera (blindness) bodeguero (winemaker) *tegué*	enceguecer (to blind) enceguecedor (blinding) *teguetá*
	/a/	plaga (plague) Moraga (a last-name) *tága*	pagar (to pay) Antofagasta (a city) *tagá*	Magallanes (a region) antofagastino (a place name) *tagatá*
	/o/	logo (logo) desahogo (relief) *kógo*	fogón (campfire) cogote (nape) *kogó*	diálogo (dialogue) catálogo (catalogue) *togotá*
	/u/	*telúgumo* *semúgu* *túgu*	mapuzugún (mapuche language) *aspugúte* *tugú*	*luguméro* *cugutérido* *tugutá*

Table 3.2. Short texts designed for the second elicitation procedure. The number of expected instances of /b/, /d/ and /g/ was balanced. Several phonetic contexts were included for each consonant. The key consonants are highlighted.

Text	Content
First	Novedades en el campo de los estudios del genoma de origen vegetal, han permitido a científicos mexicanos reproducir el código genético de una cierta variedad de planta trepadora que puede sobrevivir sin agua una cantidad sorprendente de tiempo. Según los reportes de los investigadores, la planta –pariente de la famosa Hedera helix, o hiedra común– es capaz de pasar alrededor de noventa días sin acceso alguno al vital elemento. La razón por la cual puede sobrevivir períodos de tiempo tan largos sin agua intrigó a los científicos por la mayor parte de la segunda mitad del siglo veinte; sin embargo, y gracias a la nueva tecnología disponible, recientemente se localizó un grupo de genes que parecen ser los responsables de que la planta produzca una extraña proteína que le permite sintetizar una sustancia oleosa mediante la cual se cubre completamente, evitando así la pérdida de agua por evaporación. En el fondo, la Hedera helix es una buena administradora de los recursos que tiene, lo que contrasta de forma evidente con otras estrategias que otras plantas similares han implementado.
Second	En una universidad francesa, una investigación –cuyo objetivo era conseguir resolver un misterio en torno a ciertos pagos ilegales– logró entregar pruebas contundentes para el apresamiento de la máxima autoridad de la casa de estudios superiores. Lo que ocurrió fue que el rector, Luc Mahon, realizó en agosto del año 2010 pagos por montos iguales o superiores a los setenta millones de euros a jueces de un tribunal parisino para jugar un antiguo y raro juego en el que se apuesta para predecir cómo se fallará en ciertas sentencias polémicas en casos a cargo de otros jueces. Fuentes del servicio contralor explicaron que, en situaciones como esta, el riesgo que suponen las apuestas hace que las personas se sometan a un nivel tensional que emula al de ciertas drogas estimulantes, lo que nuevamente ha supuesto volver a pensar cuáles son los nuevos y particulares problemas que las personas con altos ingresos y bien posicionadas socialmente enfrentan. ¿Cómo funciona? Es bastante complejo. Al ingresar las apuestas, los participantes agregan puntuaciones a ciertas variables probabilísticas llamadas “chances”, en rangos fijados de antemano y que además son categóricos para el tiempo total que toma el jucio (pues o se gana, o se pierde). Además, se debe calcular el “nivel de activos” que agrega

Text	Content
	un "seguro legal", figura que sigue oscuras reglas y cuyo objetivo es, para resumir, recibir las apuestas para una negociación previa en la corte; luego, se calcula primero la posibilidad real que conlleva la apuesta y, en segundo lugar, se debe asegurar que ningún activo se archive en las cortes al recibir el seguro. Evidentemente, hay que tener conocimientos avanzados del aparato legal francés y de probabilística para participar en este prestigioso e ilegal "deporte". Tras saber cómo funciona el sistema, el servicio contralor francés aseguró en un programa noticioso que el trabajo restante contempla llegar lo antes posible a conclusiones sobre cómo el rector y sus cómplices ocultaron por tanto tiempo algo tan intrincado y complejo como esto y si Mahon sabía que existían sospechas en su contra. Aparentemente, Luc Mahon debe haber tenido un amigo o colega cercano que trabaja o trabajaba para él en la corte, probablemente un abogado. De lo contrario, no se explica que siguiera en su lugar de trabajo hasta hace tan poco.

3.3.3. Recordings

Recordings were conducted by a trained phonetician (not the author) in a soundproof booth at the Phonetics Laboratory from the Pontificia Universidad Católica, in Santiago, Chile. An AT3035 cardioid condenser microphone (theoretical frequency response from 20 to 20,000 Hz, signal-to-noise ratio of 82 dB, 1 kHz at 1 Pa), a portable Mbox 2 audio interface and Protools LE 7.4.2 for Imac were used. The microphone was set approximately 15 cm from the speaker's mouth, and a pop filter was placed between the microphone and the speaker. A Hann band-stop filter from 0 to 60 Hz was applied to all recordings prior to acoustic analysis in order to remove low-frequency noise.

3.3.4. Segmentation, labelling and coding

Segmentation was conducted manually for all tokens. In the case of the word lists, only the target instances of intervocalic variants of /b/, /d/ and /g/ from words and nonsense words were segmented and labelled. In the case of short texts and semi-guided conversation, all instances of /b/, /d/ and /g/ were segmented and labelled, regardless of phonetic context. Care was taken to follow a consistent set of criteria throughout the segmentation process (Turk, Nakai & Sugahara 2006), especially given the difficulty of segmenting spirant approximant consonants, which form a continuum with their phonetic neighbours by the means of formant transitions (Martínez-Celdrán & Regueira, 2008).

In order to segment approximant realizations of /b/, /d/ or /g/, the consonant, its neighbouring segments and host word were first isolated and preliminarily segmented in TextGrids using visual cues from waveforms and spectrograms generated by *Praat*, and aided by an auditory inspection of the signals conducted in the same software (Boersma & Weenink, 2013). These preliminary boundaries for the consonant and neighbour segments were then carefully adjusted so that the approximate point of maximum constriction for the approximant consonant, as judged by visual cues and intensity contours, was located in the centre of the approximant consonant. Boundaries were finally fixed in the TextGrids inside the formant transitions, at the point where the waveform became less complex and vocalic formants started to display less energy (see Figure 3.1).

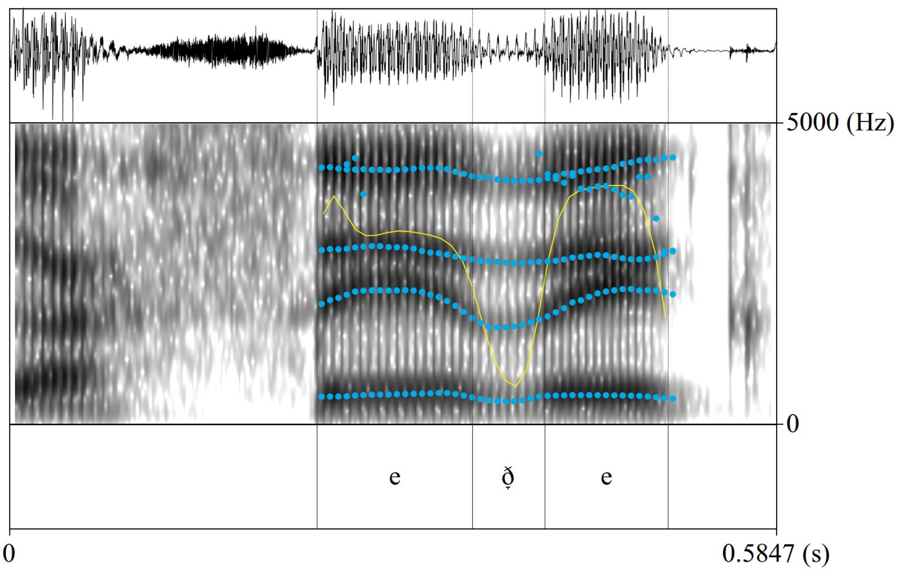

Figure 3.1. Example of a segmented open approximant consonant from the word [ˈse.ð̞e] (*sede*, “headquarters”). Formant trajectories for the first 5 oral formants are shown in blue, and the intensity contour in yellow. The scale for intensity ranges from 45 to 62 dB.

Boundaries for the segmentation of plosive realizations of /b/, /d/ and /g/ were set between the first visible silence originating from the constriction of the articulators until the end of the burst (see bottom-left panel from Figure 3.3 for an example). As for fricatives, boundaries were set at the first and last visible sections of the frication noise (see bottom-right panel from Figure 3.3 for an example).

The following variables were labelled and coded in the TextGrids: An orthographic transcription of the host word, and orthographic approximation of the actual articulation, the status of the word (word: 1; nonsense word or pause: 2), pseudo-phonetic transcription of the target and neighbouring segments, the status of the allophonic variant (vocalic approximant: 1; open approximant: 2; closed approximant: 3; plosives and fricatives: 4), and elided variants. An example of this can be seen in Figure 3.2.

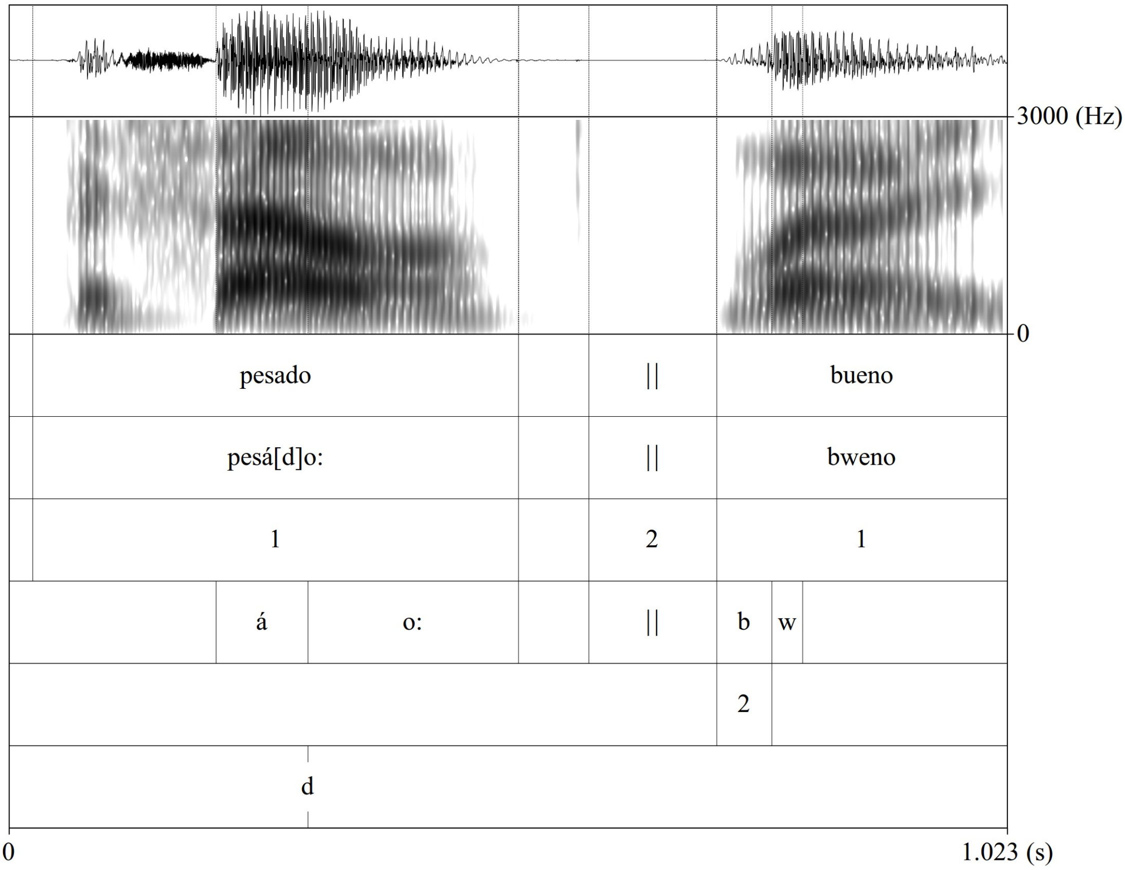

Figure 3.2. Coding example for two tokens of /d/ and /b/. As well as the waveform and spectrogram, a *Praat* TextGrid with 6 tiers is shown. From top to bottom, the tiers contain: (1) an orthographic transcription of the host words and demarcation of pauses; (2) an orthographic approximation of the realizations; (3) the word/nonsense-word status of the host words and pauses; (4) a pseudo-phonetic transcription of the target and neighbouring segments; (5) the status of the target phonetic variant pertaining type of approximant consonant or manner of articulation; and (6) elided variants.

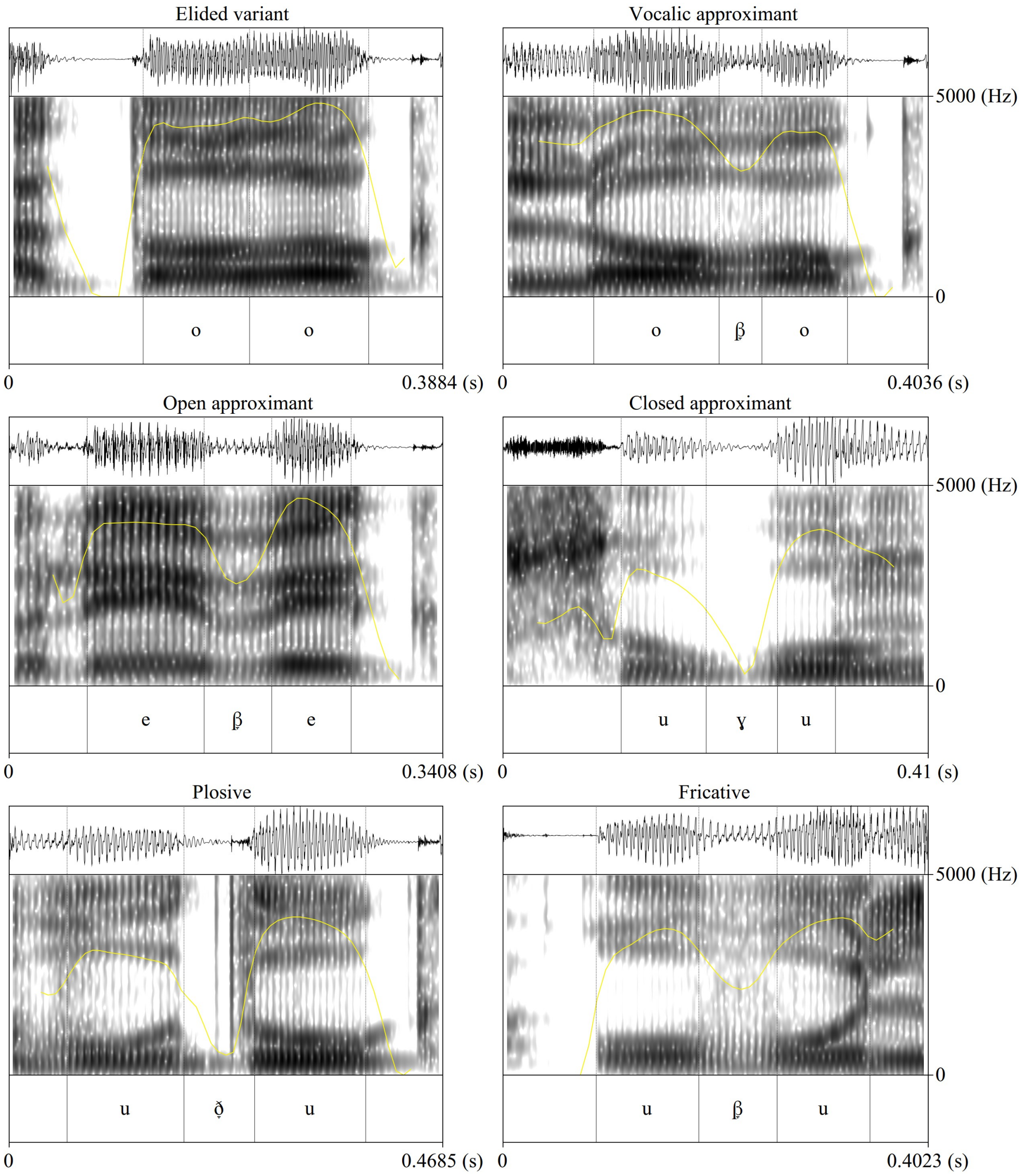

Figure 3.3. Waveform, spectrograms, intensity contours and TextGrids for the allophonic categories coded for the target segments. Top-left panel: example of an elided variant of /d/ for the word [ˈto.ð̞o] (*todo*, "all"). Top-right panel: example of a vocalic approximant of /b/ for the word [ˈlo.β̞o] (*lobo*, "wolf"). Middle-left panel: example of an open approximant of /b/ for the word [ˈbɾe.β̞e] (*breve*, "brief"). Middle-right panel: example of a closed approximant of /g/ for the word [ma.pu.su.ˈɣ̞un] (*mapuzugún*, "Mapudungun"). Lower-left panel: example of a plosive variant of /d/ for the word [bu.ˈð̞u] (*vudú*, "Vodou"). Lower-right panel: example of a fricative variant of /b/ for the word [ˈu.β̞u.la] (*úvula*, "uvula"). Intensity contours are provided on top of the spectrograms for all cases (scale 32 – 64 dB).

Spectrograms and waveforms were carefully inspected to determine the status of the allophonic variants for the target phonemes /b, d, g/. For elided variants, no visible formant movement or noticeable intensity decrease was expected between the surrounding segments. Vocalic approximants were identified as those in which a short and small formant movement and intensity decrease was visible in the signals; open approximants were expected to show longer and more pronounced formant transitions and a comparatively larger intensity decrease, resulting in the vocalic formants from F2 and upwards losing intensity; closed approximants were identified as those approximants in which there was a clear intensity decrease, resulting in formants from F2 upwards being absent, and without a characteristic plosive burst (Martínez-Celdrán, 2004, 2013; Martínez-Celdrán & Regueira, 2008). Plosives had to have a period of visible silence, VOT and burst; fricatives were required to have clear frication noise along the whole frequency scale. For examples of all these variants, see Figure 3.3.

3.3.5. Variables and measurements

The information coded in the TextGrids and files, along with a series of acoustic measurements was extracted automatically using scripts for *Praat*. In the case of intensity measurements, "Intensity" objects were created from "Sound" objects using default values. Intensity is expressed in these objects in decibels (dB), however, given that no calibration or fixed level was employed during recordings, all intensity values reported as raw dB in this thesis are in reality expressed in a scale relative to a default maximum level of 100 dB that *Praat* assigns to WAV files (in consequence, a value reported as 70 dB in this thesis actually corresponds to a value of -30 dB relative to the maximum level). As to formant measurements, "Formant" objects were created from "Sound" objects using the Burg algorithm, which performs a short-term spectral analysis to approximante spectra. Default values were used to create these objects with the exception of the maximum formant argument, which was set to 5000 Hz for recordings from male participants and to 5500 Hz for recordings from female participants. The complete list of variables can be seen in Table 3.3.

Table 3.3. List of variables extracted for each token, grouped by category.

Category	Variable	Description
Indexical and linguistic properties	Participant	Participant's identification number.
	Sex	Participant's sex.
	Task	Task corresponds to word lists, short texts or semi-guided conversation.
	Position	Relative position of target consonant within task, based on position of interval within TextGrid.
	Label	Orthographic transcription of the word hosting the target consonant.
	Label articulation	Orthographic approximation of the actual articulation for host word.
	Word boundary	Binary variable to flag tokens located next to a word boundary.
	Previous segment	Label and code of the segment preceding the target token.
	Phoneme label	Target consonant's phoneme category (/b/, /d/ or /g/).
	Following segment	Label and code of the segment following the target token.
	Variant	Allophonic variant of the target token (elided, vocalic approximant, open approximant, closed approximant or occlusives and fricatives).
	Word status	Binary variable to distinguish words from nonsense words.
Acoustic measurements	Previous segment's duration	Duration of the segment preceding the target consonant, measured in seconds.
	Target segment's duration	Target segment's duration, measured in seconds. Elided variants were entered as NA.
	Following segment's duration	Duration of the segment following the target segment, measured in seconds.
	Mean intensity (dB)	Mean intensity measured at the internal 50% duration of the target token. Elided variants were entered as NA.
	Mean F1 (Hz)	Mean F1 measured at the internal 50% duration of the target token. Elided variants were entered as NA.

Category	Variable	Description
	Mean F2 (Hz)	Mean F2 measured at the internal 50% duration of the target token. Elided variants were entered as NA.
	Previous segment's maximum intensity (dB)	Intensity at the point of maximum intensity of the segment preceding the target.
	Target segment's minimum intensity (dB)	Intensity at the point of minimum intensity of the target segment. Elided variants were entered as NA.
	Following segment's maximum intensity (dB)	Intensity at the point of maximum intensity of the segment following the target.

3.4. Normalization procedures

Normalization procedures were applied to duration, intensity and formant (F1, F2) values in order to minimize the effect of several confounding variables potentially affecting them. While most statistical analyses were conducted on the normalized data, raw acoustic values are reported for reference purposes.

3.4.1. Duration

Articulation rate has been shown to be affected by several variables such as dialect, age (Jacewicz, Fox, O'Neill & Salmons, 2009), gender (Jacewicz, Fox & Wei, 2010), elicitation tasks (Barik, 1977), and to vary within and between speakers (Jacewicz, Fox & Wei, 2010). In order to normalize for speech rate, the usual approach has been to equalize duration by dividing each measurement by a local speech rate, which can be defined operationally in various ways, but normally as words per minute (e.g., Yuan, Liberman & Cieri, 2006) or syllables per second (e.g., Fosler-Lussier & Morgan, 1999; Jacewicz, Fox, O'Neill & Salmons, 2009). An alternative is to operate under the assumption that local speech rates affect target and neighbouring segments similarly. If this assumption holds true, then duration values can be normalized by calculating the relative duration of the target consonant with respect to the segments surrounding it

(Martínez Celdrán, 2013). For this dataset, duration was normalized this way by assigning a 100% duration space to the approximant consonant and its surrounding segments, and then calculating the percentage of that duration space that the approximant consonant occupied.

Table 3.4. Skewness and excess kurtosis values from raw and normalized duration distributions of the approximant consonant variants (vocalic, open and closed approximants). Skewness and excess kurtosis differences after normalization are presented in the last column.

Approximant	**Statistic**	**Raw duration**	**Normalized duration**	**Difference**
Vocalic	Skewness	0.546	0.022	-0.524
	Excess kurtosis	0.537	0.290	-0.248
Open	Skewness	0.594	0.100	-0.494
	Excess kurtosis	1.289	0.880	-0.409
Closed	Skewness	0.100	-0.144	-0.244
	Excess kurtosis	0.880	0.085	-0.795

As shown in Figure 3.4, normalizing duration did not seem to affect the source of variation originating from the differences in the approximant consonant subcategories. Moreover, normalization reduced some of the positive skew and kurtosis in the distributions for vocalic and open approximants (see Table 3.4).

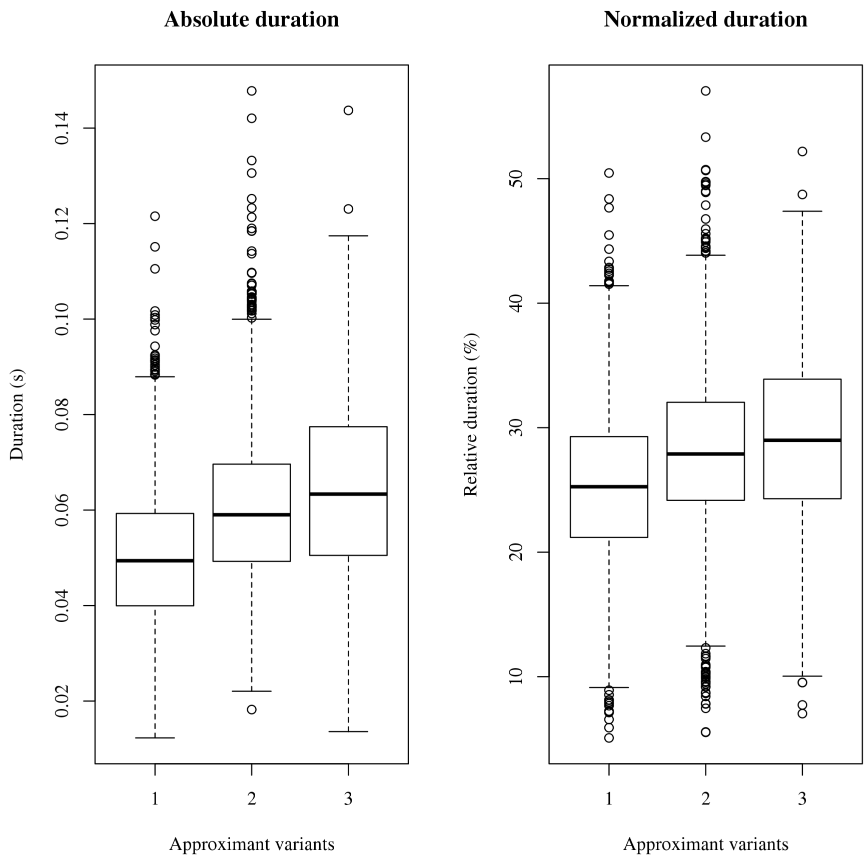

Figure 3.4. Boxplots comparing raw and normalized duration values, separately for approximant consonant variants, aggregated for /b d g/. The phonetic variants are: vocalic approximant (code: 1, n = 2184), open approximant (code: 2, n = 2261) and closed approximant (code: 3, n = 503).

3.4.2. Intensity

Absolute intensity measurements tend to incorporate undesired sources of variation originating from small fluctuations in the participant's distance to the microphone (Titze & Winholtz, 1993) and differences in the overall intensity between speakers (Hodge, Colton & Kelley, 2001), amongst other reasons. A common approach to removing these sources of variation is to normalize each intensity measurement by expressing its value as relative to a local reference level, although some caution is needed in interpreting their results (Hualde, Shosted & Scarpace, 2011). Five well known relative intensity indexes were calculated for the raw intensity values. The first index, *intensity ratio* (henceforth, "intRatio"), is calculated by dividing the target consonant minimum intensity by the following segment's maximum (Carrasco, Hualde & Simonet, 2012; Parrell, 2010). The second index, *intensity difference A* (“intDiffA”), is calculated by subtracting the target consonant's minimum intensity of the following segment's maximum (Hualde, Simonet, Shosted & Nadeu, 2010; Parrell, 2010). The third index, *intensity difference B* (“intDiffB”), is calculated by subtracting the target consonant's minimum intensity of the preceding segment's maximum intensity (Martínez-Celdrán & Regueira, 2008). The fourth index, called *maximum velocity* (“maxVel”), is calculated as the maximum intensity change from the first differences for 1 ms intervals located between the target consonant's minimum intensity and the following segment's maximum intensity (Hualde, Simonet, Shosted & Nadeu, 2010; Kingston, 2008; Parrell, 2010) as shown in Equations 3.1 and 3.2, and illustrated in Figure 3.5. For first differences at n intensity samples I_i:

$$\Delta I_i = I_i - I_{i-1} \quad \textbf{(3.1)}$$

maxVel was defined as:

$$MaxVel = \max_{i=1,\ldots,n} \Delta I_i \quad \textbf{(3.2)}$$

The fifth and final measurement, *minimum velocity* (“minVel”), is calculated as the minimum intensity change from the first differences for 1 ms intervals located between the target consonant's minimum intensity and the following segment's maximum (Kingston, 2008), as shown in Equation 3.3 and illustrated in Figure 3.5.

$$MinVel = \min_{i=1,\ldots,n} \Delta I_i \quad \textbf{(3.3)}$$

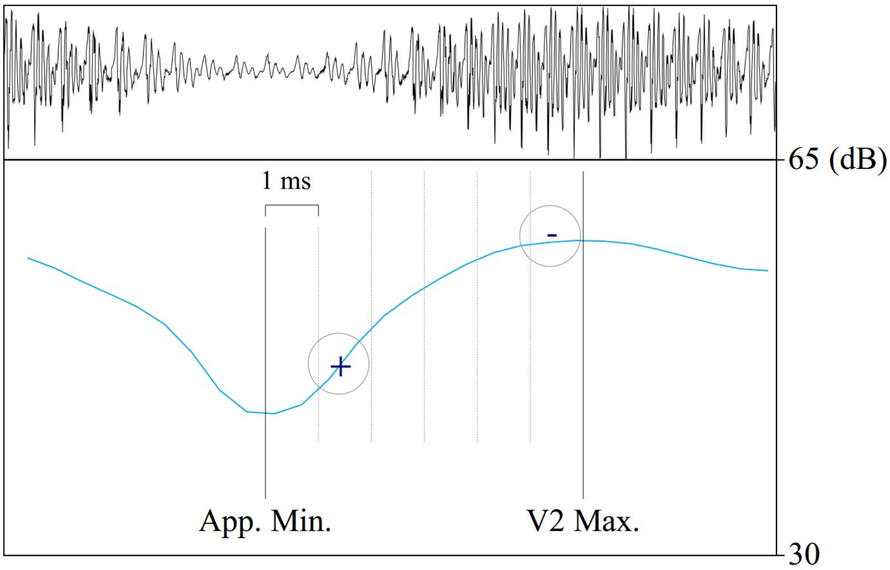

Figure 3.5. Schematic representation of maxVel and minVel calculations for intensity values. Intensity is shown as a blue contour in a scale from 30 to 65 dB. Velocity is calculated at the endpoints of 1 ms intervals equally distributed between the target consonant's minimum intensity and the following segment's maximum. Maximum and minimum velocity changes are then identified amongst all candidates, corresponding to maxVel and minVel respectively.

In order to determine the best relative intensity index for the dataset, linear discriminant analyses (LDA) and quadratic discriminant analyses (QDA) were conducted to assess the capability of each intensity index to predict the categorical variable “phonetic variant” (aggregated for /b/, /d/ and /g/), under the assumption that the levels for this variable ought to display intensity differences even after normalization, and that the relative intensity indexes should not remove this variability from the data.

As preparation for the LDA and QDA analyses, the characteristics of the distributions for the approximant variants for each intensity variable for /b, d, g/ combined were explored visually using Kernel density plots (see Figure 3.6), which display the estimated probability density function of a continuous random variable (Simonet, Hualde & Nadeu, 2012). With the exception of minVel, all intensity variables display clear differences for the approximant variant subcategories, in the directions predicted by the theoretical articulatory correlate (i.e., less constricted variants display higher intensity) or as expected by the respective transformation. With the exception of minVel, all variables display a somewhat normal distribution, although skew and positive kurtosis is clearly present for open approximant variants. Given that LDA and QDA analyses have been shown to be robust to deviations from normality (Li, Zhu & Ogihara, 2006), only minVel was discarded from further analyses at this stage. All remaining variables were z-score normalized ($\mu = 0$ and $\sigma = 1$) prior to the analyses.

For each LDA and QDA analysis, an intensity measurement or index was defined as the predictor and phonetic variant was set as the predicted dependent variable (following Adank, Smits & van Hout, 2004). The resulting variate of each analysis was used to predict the categorical variable "approximant variant" using the intensity measurements. The outcomes of these predictions were compared to the original data from the dependent variable, and percentages of correct classification were calculated for each intensity variable (see Table 3.5).

Table 3.5. Percentages of correct classification for raw intensity measurements and relative intensity indexes, for both LDA and QDA analyses. Mean intensity and minimum intensity were included as baseline predictors. While all constriction degree intensity correlates performed better than baseline, intRatio turned out to be the best predictor of approximant consonant subcategories.

Intensity variable	**LDA**	**QDA**
Mean intensity	63.6%	64.2%
Minimum intensity	65.8%	65.5%
Intensity Ratio (intRatio)	69.4%	68.8%
Intensity Difference A (intDiffA)	68.5%	68.4%
Intensity Difference B (intDiffB)	67.9%	67.5%
Maximum Velocity (maxVel)	67.9%	66.9%

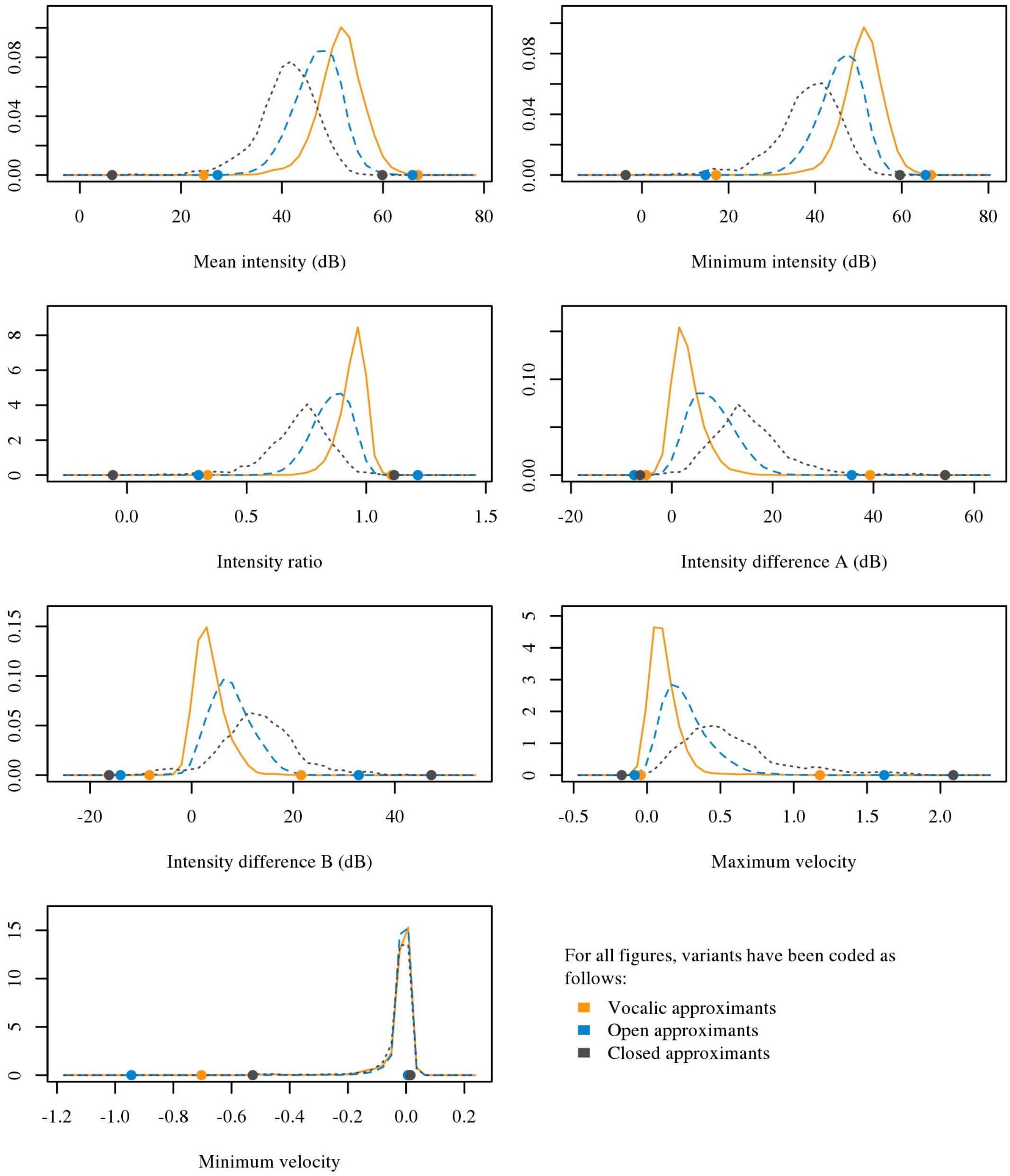

Figure 3.6. Kernel density plots for raw intensity measurements and relative intensity indexes. Density is shown as a function of acoustic measurement. Subcategories of approximant consonants are shown separately for vocalic (*n* = 2184), open (*n* = 2261) and closed approximants (*n* = 503). Coloured dots indicate the location of maximum and minimum values for each distribution in the abscissae.

The results of the LDA and QDA analyses showed that minimum intensity (LDA = 65.8%, QDA = 65.5%) was better at classifying the approximant consonants subcategories than mean intensity (LDA = 63.6%, QDA = 64.2%). This is an expected result given that mean intensity smooths out intensity differences observed in minimum intensity measurements by averaging the values around this point.

The results also showed that all relative intensity indexes performed better than baseline measurements at predicting the response variable. Consequently, it can be assumed that the normalization processes were either successful at removing sources of undesired variation from the data, or at maximizing the desired sources of variation. Amongst the intensity indexes, intRatio was the best predictor of approximant subcategories (LDA = 69.4%, QDA = 68.8%), and thus was selected to represent intensity measurements in subsequent statistical analyses.

3.4.3. Formants

It has been well established that physiological differences between speakers can have an effect on vocalic formant measurements (Peterson & Barney, 1952). Several normalization procedures have been devised to remove this variation whilst maintaining the variability due to linguistically relevant sources, such as phonological differences. However, it is not clear whether or not these procedures work equally well with approximants. Four normalization methods were assessed; Lobanov's Z-score transformation (1971), a vowel-extrinsic method; Nearey 1 (Nearey, 1977), a vowel extrinsic and formant intrinsic method, in which a separate grand mean is calculated for each formant; Nearey 2, a single logmean, vowel extrinsic, formant extrinsic approach (Nearey, 1977); and Labov's modification of Nearey's general normalization procedure, adapted for the Atlas of North American English (Labov, Ash & Boberg, 2006), a speaker-extrinsic method that employs a log-mean to achieve normalization, along with a single grand mean for all speakers. For the implementation of the normalization procedures, the package *vowels* (Kendall & Thomas, 2014) was used in *R* (R Core Team, 2013). Non-approximant variants of /b d g/ (i.e., elided, plosives and fricatives) were excluded from the normalization procedures and statistical analyses.

LDA and QDA analyses were conducted to assess the capability of the normalized data from each normalization procedure to predict the levels of two known categorical variables: *phoneme category* (levels: /b/, /d/ and /g/) and *sex* (levels: male and female). More efficient normalization methods are expected to minimize the variability present in the formant measurements originating from sex, while retaining or even improving the variability due to phonemic differences. As preparation for the analyses, the characteristics of raw and normalized distributions of F1 and F2 values for sex (Figure 3.7) and phoneme category (Figure 3.8) were explored visually using Kernel density plots. As shown in Figure 3.7, F1 and F2 formant distributions for both levels of sex appear fairly normally distributed for all variables, despite some skew being visible in some variables. Visual inspection also shows that formant differences due to sex were minimized for all normalization procedures.

As for phoneme category, F1 values look normal (see Figure 3.8), despite some kurtosis and skewness. In the case of F2, data from /b/ and /d/ display reasonably normal distributions, whereas /g/ shows a multi-modal distribution. The variability due to phoneme category differences seems to have been maintained in the data after the normalization. Given that LDA analyses have been shown to be robust to deviations from normality (Li, Zhu & Ogihara, 2006), and that the assessment of the normalization methods requires the three consonants to be present, no levels of phoneme category were discarded. All variables were centred and scaled ($\mu = 0$ and $\sigma = 1$) prior to the analyses.

LDA and QDA analyses for sex were set up with F1 and F2 as predictors and sex as the dependent variable. In the case of the analyses for phoneme category, F1 and F2 values were again defined as predictors and phoneme category was set as the predicted dependent variable. No correlations of importance were found between the predictors of any analysis ($r < |0.013|$, $p < 0.001$), and all values of the variance inflation factor (VIF) were well below the conventional threshold of 10 (VIF < 1.02; VIF $\bar{x} = 1.01$).

The resulting variates from each analysis were used to predict the levels of the dependent categorical variables, using the formant measurements as predictors. The results were then compared to the original data from the dependent variables, and a percentage of correct classification was calculated for each variable (see Table 3.6).

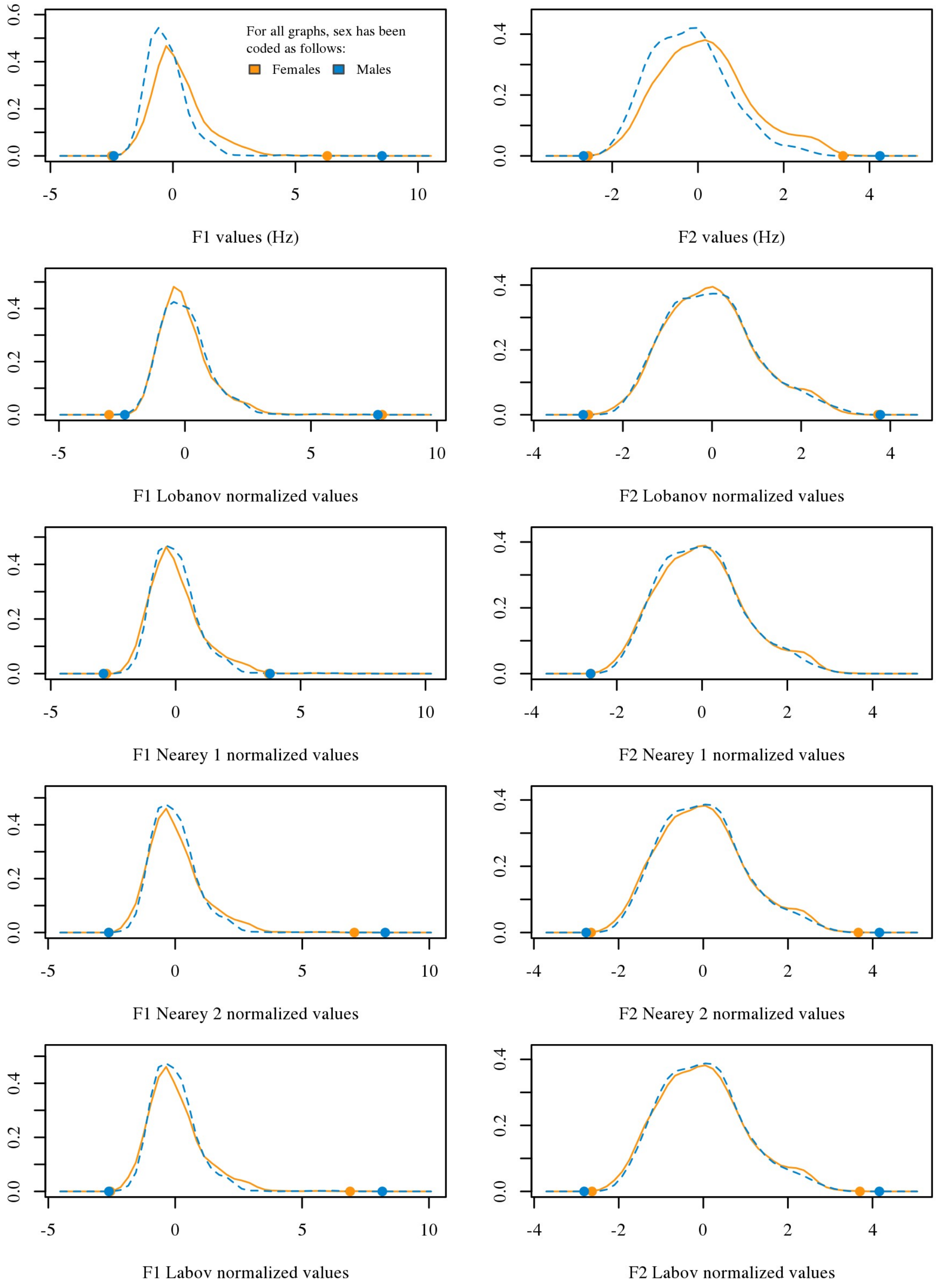

Figure 3.7. Kernel density plots for raw and normalized formant measurements, shown separately for data from female (n = 2538) and male tokens (n = 2410). Density is shown as a function of acoustic measurement or normalization result. Coloured dots indicate the location of maximum and minimum values on the horizontal axis. Formant values have been centred and scaled for all variables ($\bar{x} = 0$ and $\sigma = 1$).

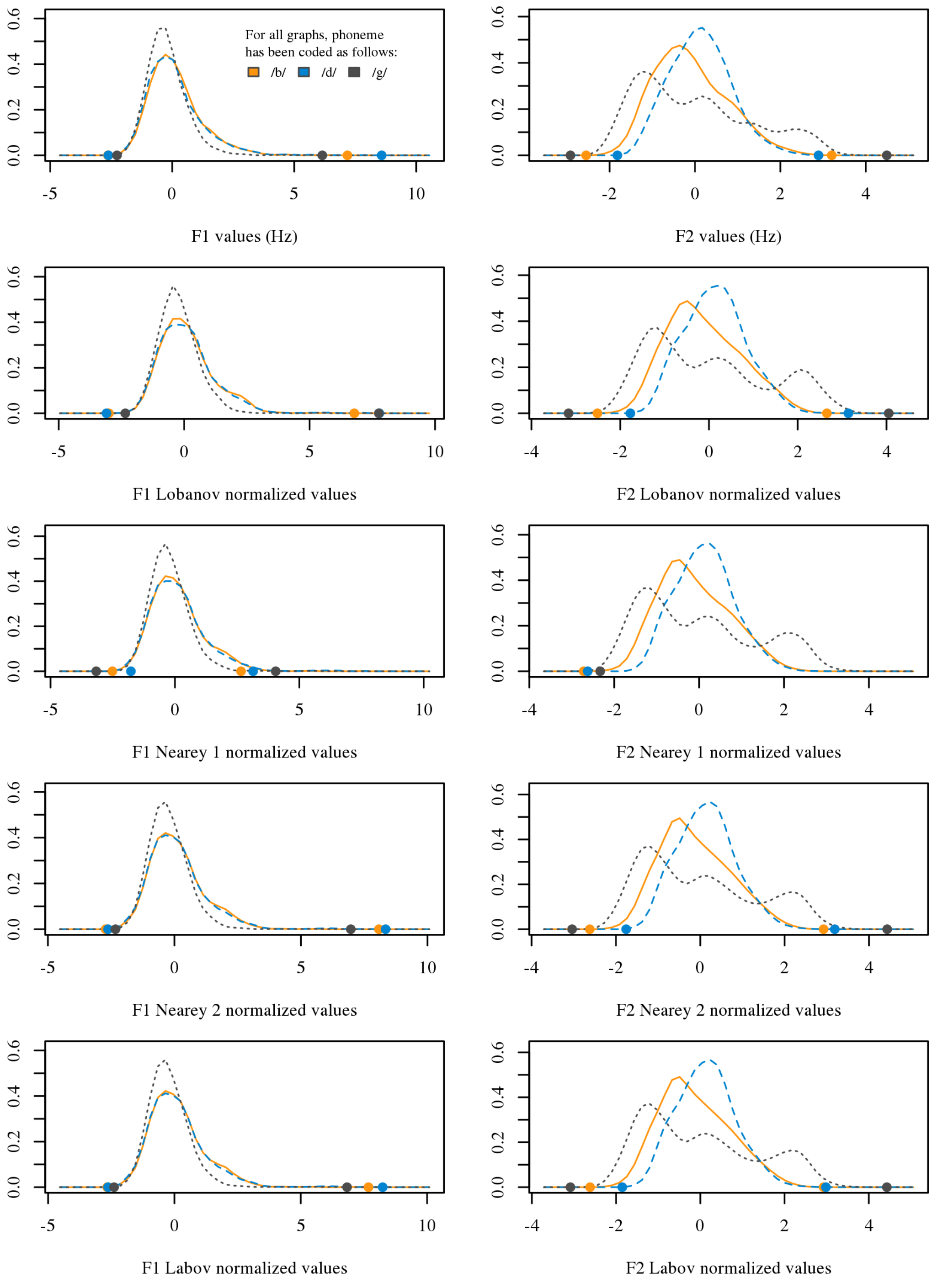

Figure 3.8. Kernel density plots for raw and normalized formant measurements, shown separately for three phoneme categories: /b/ (n = 1757), /d/ (n = 1541) and /g/ (n = 1650). Density is shown as a function of acoustic measurement or normalization result. Coloured dots indicate the location of maximum and minimum values on the horizontal axis. Formant values have been centred and scaled for all variables ($\bar{x}$ = 0 and σ = 1).

Table 3.6. Percentages of correct classification of sex and phoneme categories from raw formant measurements and normalized formant values, for both LDA and QDA analyses.

Normalization method	LDA Sex	QDA Sex	LDA Phoneme	QDA Phoneme
Mean F1-F2 (reference)	63.3%	63.7%	36.0%	48.9%
Lobanov	51.3%	51.2%	35.7%	49.8%
Nearey 1	49.4%	54.1%	35.6%	49.9%
Nearey 2	48.7%	54.3%	35.3%	49.9%
Labov's	48.6%	54.4%	35.4%	49.9%

Results of the LDA analyses for sex show that normalization procedures decreased the amount of correct classification with respect to the reference raw values, approaching chance level (50%), which is indicative of the success of the normalization procedures at removing variability originating from sex differences between participants. The results of the QDA analyses for sex replicate those from the LDA analyses.

The LDA classifications for phoneme, on the other hand, show no noticeable differences between the baseline and the normalization methods, which indicates that the normalization procedures did not remove the variability originating from phonemic differences, although results in general are all closer to chance level (33%) for LDA. Similarly, QDA analyses did not alter the accuracy of classification of phoneme categories. These analyses were considerably better than LDA analyses at classifying the levels of phoneme category (values around 50%, for a chance level of 33%). Although the differences in the results of the LDA and QDA analyses for normalized datasets are small, Lobanov normalized data performed better in QDA sex classification, and thus was selected for use in subsequent statistical analyses.

Overall, the results of the analyses confirm previous findings showing that Lobanov is an adequate normalization method for vowels (Adank, Smits & van Hout, 2004). More generally, these results showed that normalization procedures can also be successfully applied to vocoids such as spirant approximant consonants.

Chapter 4

Production of /b d g/: Results, analyses and discussion

The previous chapter described the collection of a corpus of variants of Chilean Spanish /b/, /d/ and /g/, produced in several elicitation tasks and phonetic contexts. In this chapter, the results of analyses aiming to determine the statistical significance of the acoustic differences between phonological categories and between phonetic variants within phonological categories are presented. The relationship between the variability observed in the allophones from /b d g/ and four independent categorical variables is explored as well (i.e., phonetic context, word status, elicitation procedure and lexical frequency).

4.1. Database overview

The database included 8339 instances of /b d g/. Of these, 32.9% were instances of /b/ (n = 2740), 42.6% of /d/ (n = 3556) and 24.5% of /g/ (n = 2043). As to phonetic variants, 27% of all tokens were elided (n = 2293), 26.2% corresponded to vocalic approximants (n = 2184), 27.1% to open approximants (n = 2261), 6% to closed approximants (n = 503) and 13.2% to plosives or fricatives (n = 1098). A cross-tabulation between task and phonemic category showed that, in line with expectations, items from word lists displayed similar percentages of each phonological unit, but in the other two tasks /d/ was more frequent, followed by /b/ (see Table 4.1).

There were 679 unique phonetic contexts in the database, however, most tokens were concentrated in a relatively small number of contexts, as shown by the fact that the first 100 phonetic contexts ranked by frequency accounted for 72.8% of the data (n = 6072). In order to manage this large number of contexts in subsequent analyses, they were collapsed into 9 levels, as shown in Table 4.2.

Table 4.1. Cross-tabulation of task against phoneme, displaying number of tokens on each cell, and row percentages below, in parenthesis.

Task	/b/	/d/	/g/	Row Total
Word lists	900 (33%)	921 (34%)	897 (33%)	2718 (33%)
Texts	718 (33%)	889 (41%)	588 (27%)	2195 (26%)
Conversation	1122 (33%)	1746 (51%)	558 (16%)	3426 (41%)
Column total	2740	3556	2043	8339

Table 4.2. Collapsed phonetic context categories for /b d g/ combined.

Phonetic context	Frequency	Percentage
Intervocalic	2024	24.3%
Intervocalic preceding stress	1901	22.8%
Intervocalic following stress	1843	22.1%
Following nasal	447	5.4%
Following lateral	188	2.3%
Preceding consonant	186	2.2%
Following other consonant	168	2.0%
Following pause	135	1.6%
Other	1447	17.4%

The percentage of phonetic variants for /b d g/ is detailed in Table 4.3. The largest frequencies for /b/ are open and vocalic approximants. There is a large percentage of elided variants; 23.9% of all instances (n = 654). Although there were instances of both bilabial and labiodental realisations of /b/, this articulatory distinction will not be discussed further in this chapter (see instead Chapter 5). Elided variants were the most frequent for /d/ (40.6%, n = 1445), with nearly double the instances of the next most frequent category. For /g/, the phonetic variant with the largest percentage was open approximants (41.9%, n = 856), followed by a large margin by vocalic approximants.

As opposed to /b/ and /d/, /g/ displays a smaller proportion of lenited and elided variants.

Table 4.3. Frequency and percentage of phonetic variants for /b/, /d/ and /g/. Percentages are shown below each frequency in parentheses.

Variant	**/b/**	**/d/**	**/g/**
Elided	654 (23.9%)	1445 (40.6%)	194 (9.5%)
Vocalic approximant	791 (28.9%)	795 (22.4%)	598 (29.3%)
Open approximant	825 (30.1%)	580 (16.3%)	856 (41.9%)
Closed approximant	141 (5.1%)	166 (4.7%)	196 (9.6%)
Plosives and fricatives	329 (12.0%)	570 (16.0%)	199 (9.7%)

4.2. Acoustic properties of approximant variants of /b d g/

The acoustic properties of duration, intensity and formants (F1 and F2) will be presented for the approximant variants from /b/, /d/ and /g/, aggregated for the three elicitation tasks. Descriptives and reference values will be provided for the non-normalized acoustic properties of the consonants, along with the results from statistical analyses on the normalized acoustic variables, aimed to explore the statistical significance of the differences observed between the phonological categories, differences between phonetic variants, and their interaction. In all analyses, only the acoustic properties of vocalic, open and closed approximant variants will be examined. Elided variants were excluded since these variants do not manifest acoustically; plosives and fricatives were also excluded given that their acoustic characteristics are not the focus of this study.

4.2.1. Duration

Descriptives were generated for non-normalized duration measurements from /b d g/ in order to compare these results to previous studies (see Table 4.4). In the case of /b/, vocalic approximants have a mean duration of 48 ms, open approximants 59 ms and closed approximants 65 ms. For /d/, vocalic approximants had a mean duration of 51 ms, and open and closed approximants a mean duration of 60 ms. Finally, for /g/, vocalic approximants had a mean duration of 53 ms, open approximants 61 ms and closed approximants 66 ms. Boxplots were generated for the normalized duration values for /b d g/ (see Figure 4.1); overall, the descriptives from the non-normalized measurements and the boxplots of the normalized duration values show that vocalic approximants are shorter than open approximants, and open approximants shorter than closed approximants.

Table 4.4. Descriptives for non-normalized duration values for the approximant variants of /b d g/ (“VA”: vocalic approximant; “OA”: open approximant; “CA”: closed approximant).

	/b/			**/d/**			**/g/**		
	VA	**OA**	**CA**	**VA**	**OA**	**CA**	**VA**	**OA**	**CA**
n	791	825	141	795	580	166	589	856	196
Minimum (ms)	12	23	16	20	18	19	18	23	14
Median (ms)	48	59	64	49	59	58	51	60	66
Mean (ms)	48	59	65	51	60	60	53	61	66
Maximum (ms)	92	118	144	102	133	103	122	148	112
Standard deviation (ms)	13	15	22	14	16	17	16	17	20
Skewness	0.34	0.39	0.31	0.50	0.45	0.12	0.66	0.72	-0.09
Excess kurtosis	0.01	0.59	0.38	0.21	1.14	-0.24	0.78	1.38	-0.32

A linear mixed-effects model (LMM) analysis was conducted on the normalized duration values from /b d g/ (aggregated). The analysis was carried out in *R* (R Core Team, 2013) using the *lmer* function from the *lmerTest* package (Kuznetsova, Bruun Brockhoff & Haubo Bojesen Christensen, 2015). For model selection, a null baseline model was fitted with normalized duration values from /b d g/ as the dependent variable and participant as a random factor. This model was then compared to alternative more complex models which included main effects and random slopes using the *anova* function[10], until the best fitting model was found as judged by lower Akaike information criterion values (AIC; Akaike, 1973), lower Bayesian information criterion values (BIC; Schwarz, 1978), and the statistical significance of the differences observed between the compared models, provided by a chi-squared analysis on the residuals. The best fitting model included normalized duration as the dependent variable, phonetic variant and phoneme as a main effects, their interaction, subject as a random factor, and variant as a random slope. The assumption of normality for the residuals from this model was assessed using histograms and quantile-quantile plots. No significant deviations from normality were found.

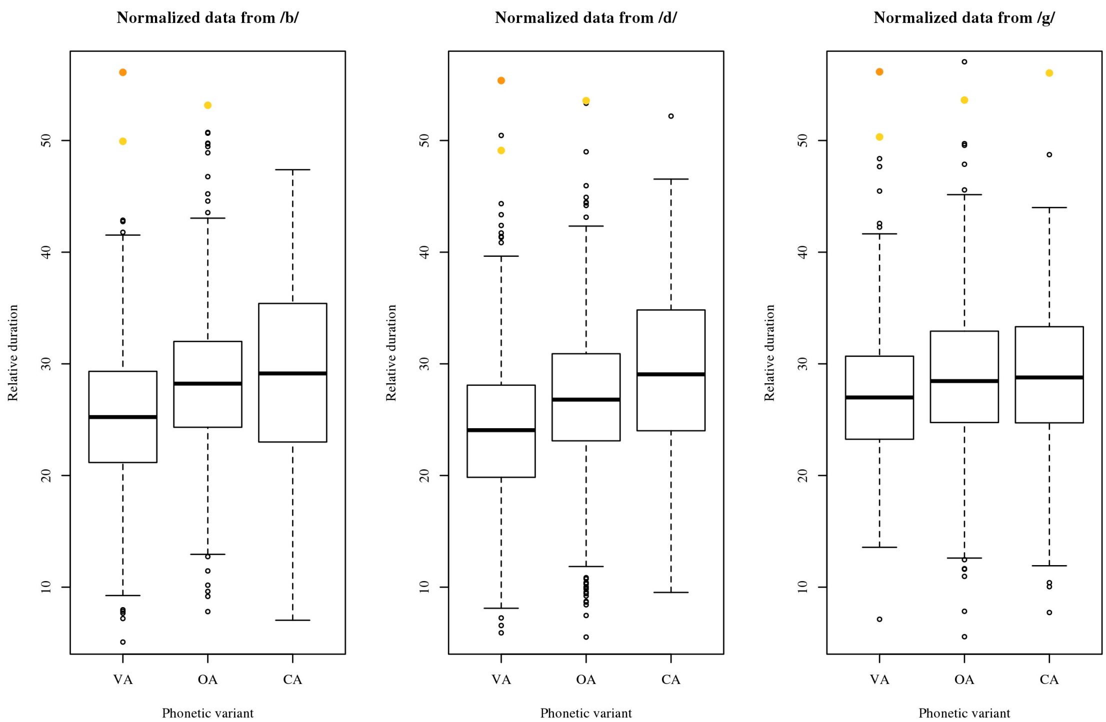

Figure 4.1. Boxplots for normalized duration values for three approximant variants of /b/, /d/ and /g/. Distances of 4 and 5 standard deviations from the median are shown as yellow and orange dots respectively.

10 In *R*, *anova* and *Anova* are different functions.

The *Anova* function from the *car* package (Fox & Weisberg, 2011) was then used to generate an analysis of deviance table (Type III) for the mixed-effects model, in order to obtain significance values for the main effects and the interaction by the means of a Wald Chi square test. A significant main effect of phoneme on the normalized durations for /b d g/ was found ($\chi^2(2) = 69.334$, $p < 0.001$), as well as a significant main effect of phonetic variant ($\chi^2(2) = 55.352$, $p < 0.001$), and a significant interaction between phonemic category and phonetic variant ($\chi^2(4) = 18.780$, $p < 0.001$). To further explore the main effect of phoneme category, the *summary* function was used on the best-fitting model to obtain statistical significance estimations for the differences observed in the response variable (normalized duration) for the levels of the independent variable phonemic category. The estimates for the *t*-tests were based on Satterthwaite approximations to degrees of freedom (Satterthwaite, 1946), and thus the *p* values should be interpreted cautiously. The comparison between the normalized values from /b/ ($\bar{x} = 26.992$, $\sigma = 6.601$) and /d/ ($\bar{x} = 25.597$, $\sigma = 6.823$) showed statistical significant differences, $t(4904) = -4.212$, $p < 0.001$, as well as the comparison between /b/ and /g/ ($\bar{x} = 28.146$, $\sigma = 6.261$), $t(4865) = 4.465$, $p < 0.001$, and the comparison between /d/ and /g/, $t(4651) = -8.317$, $p < 0.001$[11]. The general procedure applied here will be replicated for all subsequent LMM analyses, and thus only details particular to each analysis will be described in the following sections.

In order to explore the significance of normalized duration differences between phonetic variants within consonants, separate LMM analyses were conducted for /b/, /d/ and /g/. For these three analyses, best fitting models included normalized duration as a dependent variable, phonetic variant as a main effect, subject as a random factor, and variant as a random slope (no important deviations from normality were found in the residuals). A significant main effect of phonetic variant on the normalized duration for /b/ was found ($\chi^2(2) = 59.372$, $p < 0.001$), as well as significant differences between vocalic ($\bar{x} = 25.151$, $\sigma = 6.175$) and open approximants ($\bar{x} = 28.403$, $\sigma = 6.233$), $t(10.1740) = 7.273$, $p < 0.001$, and between vocalic and closed approximants ($\bar{x} = 29.067$, $\sigma = 8.213$), $t(9.2060) = 2.847$, $p < 0.05$. The comparison between open and closed approximants did not show significant differences, $t(9.3500) = 0.613$, $p = 1.11$.

11 Here and elsewhere, significance values from multiple comparisons have been corrected using the Bonferroni method (Dunn, 1959, 1961).

For /d/, a significant main effect of phonetic variant on normalized duration was found ($\chi^2(2) = 25.541$, $p < 0.001$). Significant differences were also found between vocalic ($\bar{x} = 23.961$, $\sigma = 6.263$) and open approximants ($\bar{x} = 26.882$, $\sigma = 6.694$), $t(8.7170) = 4.907$, $p < 0.01$, as well as between vocalic and closed approximants ($\bar{x}$ = 28.943 , $\sigma = 28.943$), $t(8.1910) = 3.924$, $p < 0.01$. Once more, the comparison between open and closed approximants failed to show statistically significant differences: $t(8.6260) = -1.705$, $p = 0.2476$. Finally, for /g/, a significant main effect of phonetic variant on the normalized duration values was found ($\chi^2(2) = 26.714$, $p < 0.001$). A significant difference was also found between the normalized duration values for vocalic ($\bar{x} = 26.92$, $\sigma = 5.833$) and open approximants ($\bar{x} = 28.82$, $\sigma = 6.291$): $t(6.893)$ = 4.373, $p < 0.01$. No significant differences were found between vocalic and closed approximants ($\bar{x} = 28.949$, $\sigma = 6.816$), $t(8.351) = 0.953$, $p = 0.734$, nor between open and closed approximants, $t(7.861) = 0.545$, $p = 1.202$.

4.2.2. Intensity

Results for /b/ showed that vocalic approximants had a mean minimum intensity of 51.4 dB, open approximants 46.5 dB and closed approximants 39.6 dB (see Table 4.5). In the case of /d/, vocalic approximants had a mean minimum intensity of 51.3 dB, open approximants 45.4 dB and closed approximants 36.2 dB. Finally, for /g/, vocalic approximants had a mean minimum intensity of 50.1 dB, open approximants 45.7 dB and closed approximants 38.9 dB. Boxplots were generated for the normalized intensity values (intensity ratio) for the phonetic variants of /b/, /d/ and /g/ (see Figure 4.2). Overall, vocalic approximants displayed higher intensity ratio values, followed by open and then closed approximants. Several outliers can be seen in most distributions, resulting in negatively skewed distributions, although this has no negative impact on LMM analyses.

A LMM analysis was conducted on the intensity ratio values from /b d g/ (aggregated). The best fitting model included intensity ratio as the dependent variable, phonetic variant and phoneme category as a main effects, their interaction, subject as a random factor, and variant as a random slope (no important deviations from normality were found in the residuals from this analysis). A significant main effect of phoneme category on intensity ratio was found ($\chi^2(2) = 61.048$, $p < 0.001$), as well as a significant main effect of phonetic variant ($\chi^2(2) = 265.438$, $p < 0.001$), and a significant interaction between phonemic category and phonetic variant ($\chi^2(4) = 104.756$, $p < 0.001$). Post hoc *t*-tests obtained through the *summary* function revealed significant differences in intensity ratio values between /b/ ($\bar{x} = 0.896$, $\sigma = 0.091$) and /g/ ($\bar{x} = 0.860$, $\sigma = 0.101$), $t(4737) = -6.588$, $p < 0.001$, and between /d/ ($\bar{x} = 0.881$, $\sigma = 0.116$) and /g/, $t(4269) = 7.221$, $p < 0.001$, but not between /b/ and /d/, $t(4859) = 0.708$, $p = 0.958$.

Table 4.5. Descriptives for minimum intensity measurements for approximant variants of /b/, /d/ and /g/ (“VA”: vocalic approximant; “OA”: open approximant; “CA”: closed approximant).

	/b/			**/d/**			**/g/**		
	VA	**OA**	**CA**	**VA**	**OA**	**CA**	**VA**	**OA**	**CA**
n	791	825	141	795	580	166	589	856	196
Minimum (dB)	24.8	17.5	20.9	19.4	29.8	1.2	35.9	26.5	14.9
Median (dB)	51.5	47.2	40.3	51.6	45.7	37.8	50.5	45.8	39.4
Mean (dB)	51.4	46.5	39.6	51.3	45.4	36.2	50.1	45.7	38.9
Maximum (dB)	62.1	60.1	51.5	64.5	60.3	51.8	62.4	62.6	54.6
Standard deviation (dB)	4.3	4.9	5.9	4.6	4.9	8.4	4.2	5.0	7.1
Skewness	-0.71	-0.79	-0.60	-1.03	-0.26	-1.33	-0.39	-0.35	-0.73
Excess kurtosis	2.52	1.73	0.48	4.51	0.02	2.45	0.90	0.55	0.53

Separate LMM analyses were conducted for /b/, /d/ and /g/, in order to explore the significance of intensity ratio differences between phonetic variants within phonemic category. For these analyses, best fitting models included intensity ratio as dependent variable, phonetic variant as a main effect, subject as a random factor, and variant as a random slope. The assumption of normality for the residuals was assessed using histograms and quantile-quantile plots. Although some negative skew (/b/ = -1.247; /d/ = -1.468; /g/ = -0.546) and positive excess kurtosis was observed (/b/ = 6.894; /d/ = 8.178; /g/ = 2.277), the residuals were still considered to be fairly normally distributed. In the case of /b/, a main effect of phonetic variant on the normalized intensity values was found ($\chi^2(2) = 141.55$, $p < 0.001$). Differences in the normalized intensity values for

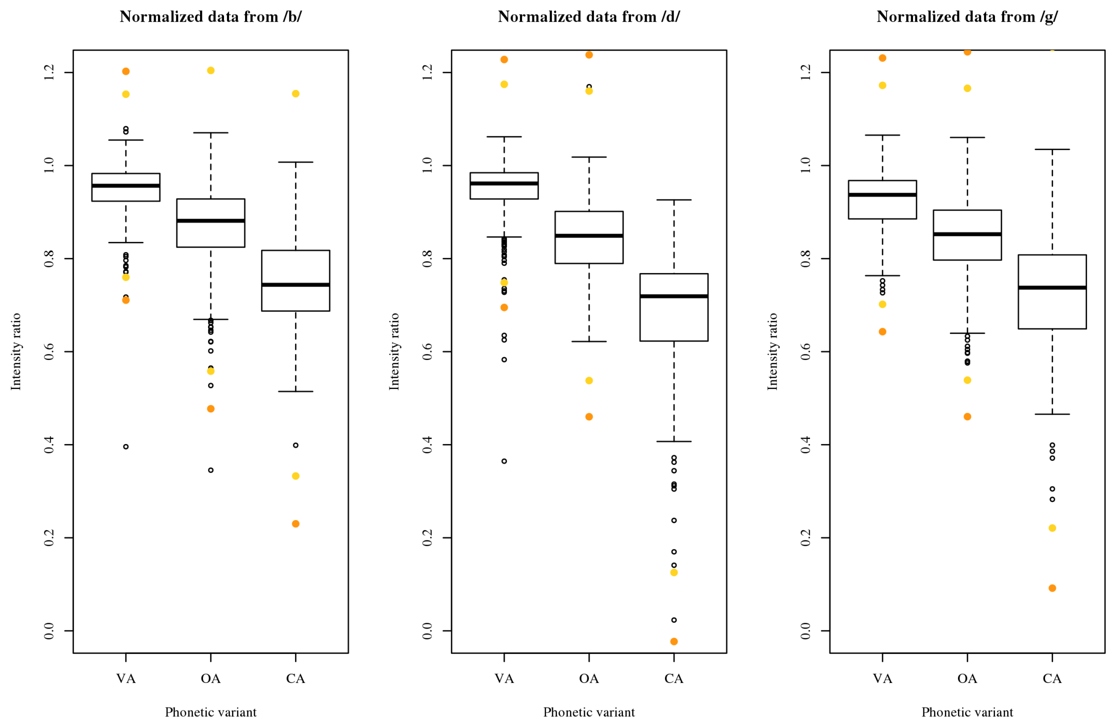

Figure 4.2. Boxplots for normalized intensity values for three approximant variants of /b/, /d/ and /g/. Distances of 4 and 5 standard deviations from the median are shown as yellow and orange dots respectively.

vocalic ($\bar{x}$ = 0.949, σ = 0.049) and open approximants ($\bar{x}$ = 0.87, σ = 0.081) were statistically significant, $t(8.535) = -10.76$, $p < 0.001$, as well as those between vocalic and closed approximants ($\bar{x}$ = 0.746, σ = 0.103), $t(7.192) = -11.73$, $p < 0.001$, and between open and closed approximants, $t(6.539) = 10.46$, $p < 0.001$. For /d/, a significant main effect of phonetic variant was found ($\chi^2(2) = 459.74$, $p < 0.001$), along significant differences between the normalized intensity values of vocalic ($\bar{x}$ = 0.951, σ = 0.053) and open approximants ($\bar{x}$ = 0.844, σ = 0.078), $t(7.44) = -16.12$, $p < 0.001$, between vocalic and closed approximants ($\bar{x}$ = 0.680, σ = 0.148), $t(8.635) = -12.47$, $p <$ 0.001, and between open and closed variants: $t(8.749) = 6.885$, $p < 0.001$. Finally, for /g/, a significant main effect of phonetic variant was also found ($\chi^2(2) = 141.55$, $p <$ 0.001). Furthermore, significant differences were found between vocalic ($\bar{x}$ = 0.924, σ = 0.059) and open approximants ($\bar{x}$ = 0.846, σ = 0.078), $t(8.38) = -9.415$, $p < 0.001$, as well as for vocalic and closed approximants ($\bar{x}$ = 0.724, σ = 0.129), $t(9.154) = -11.278$, $p < 0.001$, and between open and closed approximants: $t(9.792) = 7.242$, $p < 0.001$.

4.2.3. Formants (F1, F2)

Results for F1

Results for the non-normalized values for /b/ showed that vocalic approximants had a mean F1 value of 473.9 Hz, open approximants 417.7 Hz and closed approximants 346.3 Hz (see Table 4.6). In the case of /d/, vocalic approximants had a value of 476.9 Hz, open approximants 388.7 Hz and closed approximants 383.8 Hz. The results for /g/ showed mean values of 423.0 Hz, 382.5 Hz and 382.2 Hz for vocalic, open and closed approximants respectively. Boxplots for the Lobanov-normalized formant values for /b/, /d/ and /g/, showed that, for the three consonants, F1 values decreased as a function of phonetic variant, with vocalic approximants displaying the highest values, followed by open approximants, and then by closed approximants (see Figure 4.3). Clear signs of positive skew were observed for most distributions.

Table 4.6. Descriptives for mean F1 measurements for approximant variants of /b/, /d/ and /g/ ("VA": vocalic approximant; "OA": open approximant; "CA": closed approximant).

	/b/			**/d/**			**/g/**		
	VA	**OA**	**CA**	**VA**	**OA**	**CA**	**VA**	**OA**	**CA**
n	791	825	141	795	580	166	589	856	196
Minimum (Hz)	236.7	205.0	175.5	216.8	210.9	179.9	240.1	203.1	218.3
Median (Hz)	453.7	398.2	311.6	457.7	373.8	315.5	412.0	374.6	351.1
Mean (Hz)	473.9	417.7	346.3	476.9	388.7	383.8	423.0	382.5	382.2
Maximum (Hz)	1113.0	973.4	1211.5	859.7	1019.1	1375.9	775.2	779.4	1105.4
Standard deviation (Hz)	122.9	107.4	150.7	118.0	95.0	211.0	87.1	79.5	142.7
Skewness	0.91	1.26	2.83	0.80	1.56	2.47	0.76	0.78	2.69
Excess kurtosis	1.14	2.50	10.90	0.38	4.95	6.21	0.69	1.45	9.01

A LMM analysis was conducted on the normalized F1 values from /b d g/ (aggregated). The best fitting model included normalized F1 as the dependent variable, phonetic variant and phoneme category as a main effects, their interaction, subject as a random factor, and variant as a random slope (no important deviations from normality were found in the residuals from this analysis). A significant main effect of phoneme category on normalized F1 was found ($\chi^2(2) = 107.170$, $p < 0.001$), as well as a significant main effect of phonetic variant ($\chi^2(2) = 226.645$, $p < 0.001$), and a significant interaction between phonemic category and phonetic variant ($\chi^2(4) = 59.518$, $p < 0.001$). Post hoc *t*-tests obtained through the *summary* function revealed significant differences in normalized F1 values between /b/ ($\bar{x} = 0.112$, $\sigma = 1.037$) and /g/ ($\bar{x} = -0.229$, $\sigma = 0.792$), $t(4758) = -8.846$, $p < 0.001$, as well as between /d/ ($\bar{x} = 0.117$, $\sigma = 1.109$) and /g/, $t(4319) = 9.482$, $p < 0.001$, but not between /b/ and /d/, $t(4863) = 0.721$, $p = 0.9418$.

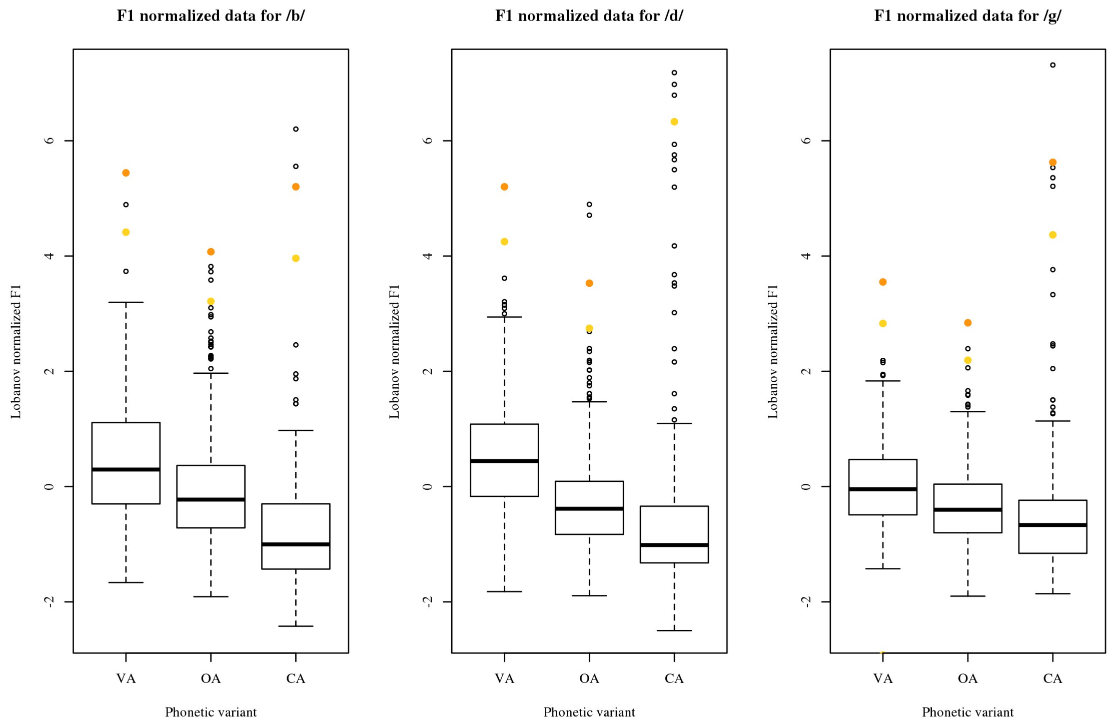

Figure 4.3. Boxplots for Lobanov-normalized F1 values for three approximant variants of /b/, /d/ and /g/. Distances of 4 and 5 standard deviations from the median are shown as yellow and orange dots respectively.

Separate LMM analyses were conducted for /b/, /d/ and /g/, in order to explore the significance of normalized F1 differences between phonetic variants within phonemic category. The best fitting model for all these analyses included normalized F1 as dependent variable, phonetic variant as a main effect and subject as a random factor.

The LMM for /d/ and /g/ also included variant as a random slope. The assumption of normality for the residuals was assessed using histograms and quantile-quantile plots. Some signs of positive skew (/b/ = 1.091; /d/ = 1.67; /g/ = 1.847) and indications of positive excess kurtosis were observed (/b/ = 2.98; /d/ = 6.882; /g/ = 11.543), and thus the following results need to be considered with caution. In the case of /b/, a main effect of phonetic variant on the normalized F1 was found: $\chi^2(2) = 257.57$, $p < 0.001$. Significant differences for the normalized F1 values were also found between vocalic ($\bar{x} = 0.473$, $\sigma = 1.03$) and open approximants ($\bar{x} = -0.101$, $\sigma = 0.86$), $t(1606.9) = -12.421$, $p < 0.001$, as well as between vocalic and closed approximants ($\bar{x} = -0.667$, $\sigma = 1.241$), $t(1741) = -13.387$, $p < 0.001$, and between open and closed approximants: $t(1753.5) = -6.521$, $p < 0.001$.

The results for /d/ showed a significant main effect of phonetic variant on normalized F1 ($\chi^2(2) = 257.29$, $p < 0.001$). Statistically significant differences were also found between vocalic ($\bar{x} = 0.517$, $\sigma = 0.952$) and open approximants ($\bar{x} = -0.289$, $\sigma = 0.783$), $t(28.311) = -13.375$, $p < 0.001$, and between vocalic and closed approximants ($\bar{x} = -0.378$, $\sigma = 1.836$), $t(8.603) = -3.941$, $p < 0.01$. The comparison between open and closed approximants did not show statistically significant differences: $t(8.409) = 0.176$, $p = 1.73$. Finally, for /g/, a significant main effect of phonetic variant was found ($\chi^2(2) = 95.341$, $p < 0.001$), as well as significant differences between vocalic ($\bar{x} = 0.018$, $\sigma = 0.719$) and open approximants ($\bar{x} = -0.354$, $\sigma = 0.649$), $t(9.978) = -5.943$, $p < 0.001$, and between vocalic and closed approximants ($\bar{x} = -0.432$, $\sigma = 1.259$), $t(8.594) = -3.563$, $p < 0.05$. No significant differences were found between open and closed approximants: $t(11.058) = 0.634$, $p = 1.078$.

Results for F2

Results for the non-normalized F2 values for /b/ showed that vocalic approximants had a mean F2 value of 1471 Hz, open approximants 1455.3 Hz and closed approximants 1461.8 Hz (see Table 4.7). In the case of /d/, vocalic approximants had a mean F2 value of 1559.4 Hz, open approximants 1574.3 Hz and closed approximants 1639.2 Hz. The results for /g/ showed that vocalic approximants had a mean value of 1592.1 Hz, open approximants 1438.6 Hz and closed approximants 1565.9 Hz (see

Table 4.15). Boxplots for the Lobanov-normalized formant values for /b/, /d/ and /g/ showed no clear pattern of F2 variation differentiating the phonetic variants within phonemic category (see Figure 4.4). In general, the distributions of the phonetic variants overlap to a significant degree. Clearer differences are observed between consonants, with the normalized results from /g/ displaying a wider range of values, followed by /b/ and then by /d/.

Table 4.7. Descriptives for mean F2 measurements for approximant variants of /b/, /d/ and /g/ ("VA": vocalic approximant; "OA": open approximant; "CA": closed approximant).

	/b/			**/d/**			**/g/**		
	VA	**OA**	**CA**	**VA**	**OA**	**CA**	**VA**	**OA**	**CA**
n	791	825	141	795	580	166	589	856	196
Minimum (Hz)	518.9	498.0	558.8	833.5	820.3	910.8	493.2	473.3	471.0
Median (Hz)	1403.7	1404.1	1466.1	1548.1	1576.7	1653.9	1549.6	1254.2	1433.2
Mean (Hz)	1471.0	1455.3	1461.8	1559.4	1574.3	1639.2	1592.1	1438.6	1565.9
Maximum (Hz)	2853.6	2816.8	2495.0	2681.4	2662.1	2740.9	3010.3	3088.1	3339.0
Standard deviation (Hz)	440.6	397.2	399.7	349.5	328.9	322.6	613.3	676.8	734.7
Skewness	0.55	0.49	0.10	0.27	0.36	0.31	0.41	0.71	0.59
Excess kurtosis	-0.26	-0.08	-0.59	-0.27	-0.03	0.53	-0.76	-0.61	-0.64

A LMM analysis was conducted on the normalized F2 values from /b d g/ (aggregated). The best fitting model included normalized F2 as the dependent variable, phonetic variant and phoneme category as a main effects, their interaction, and subject as a random factor (no important deviations from normality were found in the residuals from this analysis). A significant main effect of phoneme category on normalized F2 was found ($\chi^2(2) = 27.17$ $p < 0.001$), as well as a significant interaction between phonetic category and phonetic variant ($\chi^2(4) = 20.37$, $p < 0.001$), but there was no main effect of phonetic variant ($\chi^2(2) = 3.31$ $p = 0.1913$). Post hoc *t*-tests obtained through the *summary* function revealed significant differences in normalized F2 values between /b/ ($\bar{x} = -0.119$, $\sigma = 0.816$) and /d/ ($\bar{x} = 0.134$, $\sigma = 0.666$), $t(4939) = 3.963$, $p < 0.001$, as well as between /b/ and /g/ ($\bar{x} = 0.002$, $\sigma = 1.358$), $t(4939) = 4.840$, $p < 0.001$, but not between /d/ and /g/, $t(4939) = -1.169$, $p = 0.485$.

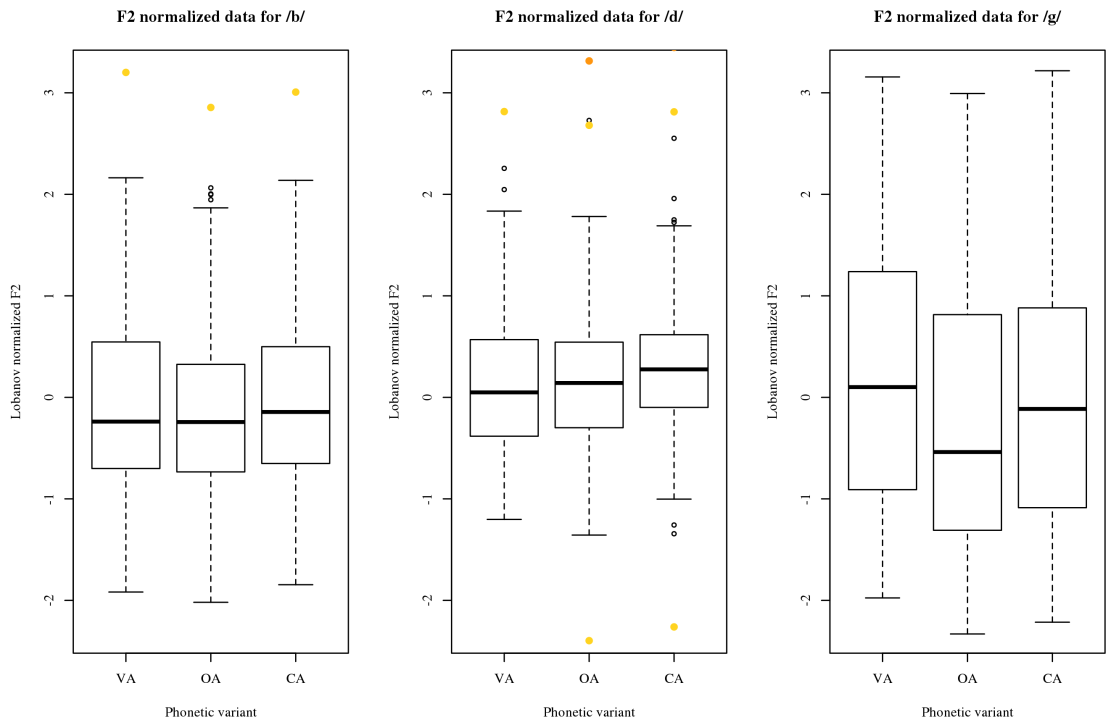

Figure 4.4. Boxplots for Lobanov-normalized F2 values for three approximant variants of /b/, /d/ and /g/. Distances of 4 and 5 standard deviations from the median are shown as yellow and orange dots respectively.

Separate LMM analyses were conducted for /b/, /d/ and /g/, in order to explore the significance of normalized F2 differences between phonetic variants within phonemic category. The best fitting model for these three analyses included normalized F2 as dependent variable, phonetic variant as a main effect and subject as a random factor (no important deviations from normality were found in the residuals from these models). No significant main effect of phonetic variant on the normalized F2 values was found for /b/ ($\chi^2(2) = 4.534$, $p = 0.104$). Moreover, the *t*-tests from the *summary* function did not reveal significant F2 differences between the phonetic variants. In the case of /d/, a significant main effect of phonetic variant was found ($\chi^2(2) = 7.5607$, $p < 0.05$), as well as statistically significant differences between vocalic ($\overline{x} = 0.115$, $\sigma = 0.692$) and closed approximants ($\overline{x} = 0.263$, $\sigma = 0.634$), $t(1522) = 2.660$, $p < 0.05$, and between open ($\overline{x} = 0.122$, $\sigma = 0.635$) and closed approximants, $t(1531.6) = -2.546$, $p < 0.05$. No such differences were found between vocalic and open approximants: $t(1468) = 0.054$, $p =$

1.9136. Finally, for /g/, a significant main effect of phonetic variant was found ($\chi^2(2) = 19.511$, $p < 0.001$), along with significant differences between vocalic ($\bar{x} = 0.178$, $\sigma = 1.242$) and open approximants ($\bar{x} = -0.136$, $\sigma = 1.407$), $t(1647) = -4.363$, $p < 0.001$, but not between vocalic and closed approximants ($\bar{x} = 0.064$, $\sigma = 1.421$), $t(1647) = -1.023$, $p = 0.613$, or between open and closed approximants: $t(1647) = -1.873$, $p = 0.123$.

4.3. The role of indexical variables in the production of /b d g/

4.3.1. Phonetic context

Most phonetic variants from /b/ surfaced intervocalically, with elided and vocalic approximants predominating (see Table 4.8). Plosive and fricative variants were more frequent after nasals.

To explore whether phonetic variant can be predicted from phonetic context, a multinomial logistic regression analysis (MLR) was conducted in *R* (R Core Team, 2013) using the *mlogit* package (Croissant, 2013). The cross-tabulation of phonetic variant against phonetic context revealed that some cells contained fewer than 10 cases (see Table 4.8), so relevelling was required (Schwab, 2002): the phonetic contexts

Table 4.8. Cross-tabulation of phonetic context against phonetic variant for variants from /b/ (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “PF”: plosives and fricatives).

Phonetic contexts	EL	VA	OA	CA	PF	Sum	(%)
Intervocalic	198	269	184	8	18	677	(24.7%)
Intervocalic preceding stress	113	190	261	37	43	644	(23.5%)
Intervocalic following stress	172	192	147	10	29	550	(20.1%)
Following nasal	23	2	3	7	80	115	(4.2%)
Preceding consonant	39	15	8	9	9	80	(2.9%)
Following lateral	0	2	33	14	15	64	(2.3%)
Following other consonant	1	5	11	5	13	35	(1.3%)
Following pause	0	0	7	5	14	26	(0.9%)
Other	108	116	171	46	108	549	(20%)
Sum	654	791	825	141	329	2740	

"following pause", "following lateral" and "following nasal" were collapsed into "following strong"; similarly, "preceding consonant" and "following other consonant" were collapsed into "other consonants". Phonetic contexts under "other" were not included in the analysis given that this category does not represent a unitary interpretable context. Phonetic variant was defined as a dependent variable (reference level: open approximant) and collapsed phonetic context as the independent predictor variable (reference level: intervocalic preceding stress). The results for this analysis are summarized in Table 4.9. The same general procedure was used in all subsequent MLR analyses.

Phonetic context was a good predictor of phonetic variant for /b/. Within each comparison, most phonetic contexts were able to significantly predict whether a variant would surface as a given level instead of the reference value "open approximant". The estimates and odds ratios were also in line with expectations, showing that it was more likely for lenited variants to surface in weaker phonetic contexts (e.g, intervocalic), while the opposite was true for less lenited variants.

Table 4.9. Summary of results for an MLR analysis for /b/, with phonetic variant as dependent variable and phonetic context as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	-0.837(0.113)	< 0.001	***			
Intervocalic	0.910(0.152)	< 0.001	***	1.844	2.485	3.349
Int. following stress	0.994(0.159)	< 0.001	***	1.979	2.703	3.691
Int. following strong	0.211(0.282)	= 0.453		0.711	1.235	2.146
Other consonants	1.582(0.301)	< 0.001	***	2.698	4.863	8.763
Vocalic vs. Open app.						
(Intercept)	-0.317(0.095)	< 0.001	***			
Intervocalic	0.697(0.135)	< 0.001	***	1.541	2.008	2.617
Int. following stress	0.585(0.145)	< 0.001	***	1.350	1.794	2.385

Int. following strong	-2.057(0.531)	< 0.001	***	0.045	0.128	0.362
Other consonants	0.369(0.334)	= 0.270		0.751	1.446	2.784
Closed vs. Open app.						
(Intercept)	-1.954(0.176)	< 0.001	***			
Intervocalic	-1.182(0.402)	< 0.01	**	0.140	0.307	0.674
Int. following stress	-0.734(0.371)	< 0.05	*	0.232	0.480	0.993
Int. following strong	1.450(0.304)	< 0.001	***	2.349	4.265	7.743
Other consonants	1.648(0.394)	< 0.001	***	2.403	5.198	11.242
Other vs. Open app.						
(Intercept)	-1.803(0.165)	< 0.001	***			
Intervocalic	-0.521(0.297)	< 0.1	.	0.332	0.594	1.062
Int. following stress	0.180(0.261)	= 0.491		0.717	1.197	1.999
Int. following strong	2.733(0.244)	< 0.001	***	9.538	15.386	24.819
Other consonants	1.950(0.354)	< 0.001	***	3.513	7.028	14.060

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood (unexplained variability) = -2873. McFadden R^2 (effect size): 0.087695. Likelihood ratio test (significant variability explained by model): $\chi^2 = 552.33$, $p < 0.001$.

As was the case for /b/, most instances of /d/ occurred in intervocalic contexts (see Table 4.10). Highly lenited variants predominated in this subset, followed by open approximants. In order to explore whether the level of phonetic variant can be predicted from phonetic context, an MLR analysis was conducted – relevelling was required – with phonetic variant as the predicted variable (reference level: open approximant) and collapsed phonetic context as the predictor variable (reference level: preceding stress). The results from this analysis are summarized in Table 4.11.

Phonetic context was a good predictor of phonetic variant for /d/. Most phonetic contexts were able to significantly predict whether a variant would surface at a given allophonic level instead of open approximant. Estimates and odds ratios were again in line with expectations, showing that it was more likely for lenited variants to surface in weaker phonetic contexts. Perhaps an exception were elided variants for /d/ following a strong context (OR = 6.025). This result might be due to 61 cases of elided variants after a nasal (see Table 4.11), which corresponded to a coarticulatory effect between nasals and the following segment, in which the former was unreleased and the following nasalized consonant lacked a visible onset (see Figure 4.5 for an example).

Table 4.10. Cross-tabulation of phonetic context against phonetic variant for variants from /d/ (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “PF”: plosives, fricatives and others).

Phonetic contexts	EL	VA	OA	CA	PF	Sum	(%)
Intervocalic	453	268	122	27	36	906	(25.5%)
Intervocalic following stress	439	222	134	18	16	829	(23.3%)
Intervocalic preceding stress	228	189	157	23	43	640	(18%)
Following nasal	61	2	2	8	193	266	(7.5%)
Following other consonant	13	9	41	39	26	128	(3.6%)
Following pause	0	1	2	11	93	107	(3%)
Preceding consonant	57	5	1	0	1	64	(1.8%)
Following lateral	9	2	4	3	36	54	(1.5%)
Other	185	97	117	37	126	562	(15.8%)
Sum	1445	795	580	166	570	3556	

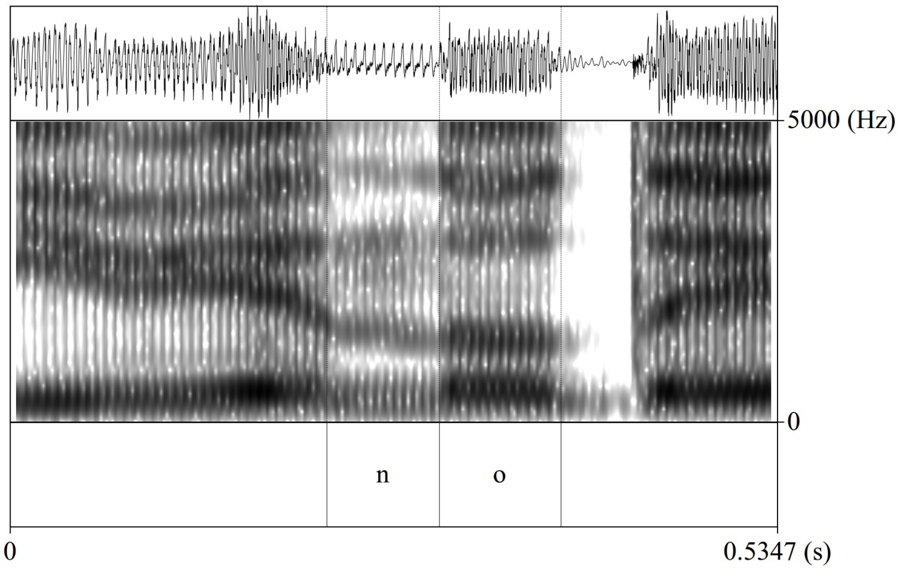

Figure 4.5. Waveform and spectrogram for the word *viendo* /ˈbien.do/, “seeing”; expected articulation [ˈbjen̪.d̪o]. No instance of /d/ is visible in the spectrogram; the small burst between the consonant and the vowel corresponds to a nasal release.

Table 4.11. Summary of results for an MLR analysis for /d/, with phonetic variant as dependent variable and phonetic context as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	0.373(0.104)	< 0.001	***			
Intervocalic	0.939(0.145)	< 0.001	***	1.923	2.557	3.400
Int. following stress	0.814(0.143)	< 0.001	***	1.704	2.256	2.987
Int. following strong	1.796(0.387)	< 0.001	***	2.820	**6.025**	12.873
Other consonants	0.138(0.221)	= 0.533		0.744	1.148	1.770
Vocalic vs. Open app.						
(Intercept)	0.186(0.108)	< 0.1	.			
Intervocalic	0.601(0.154)	< 0.001	***	1.350	1.825	2.466
Int. following stress	0.319(0.154)	< 0.05	*	1.018	1.376	1.860
Int. following strong	-0.656(0.580)	= 0.259		0.167	0.519	1.619
Other consonants	-1.284(0.327)	< 0.001	***	0.146	0.277	0.526
Closed vs. Open app.						
(Intercept)	-1.921(0.223)	< 0.001	***			
Intervocalic	0.413(0.308)	= 0.181		0.825	1.511	2.765
Int. following stress	-0.087(0.336)	= 0.796		0.475	0.917	1.771
Int. following strong	2.932(0.469)	< 0.001	***	7.481	18.772	47.101
Other consonants	1.847(0.315)	< 0.001	***	3.418	6.339	11.755
Other vs. Open app.						
(Intercept)	-1.295(0.172)	< 0.001	***			
Intervocalic	0.075(0.256)	= 0.771		0.652	1.077	1.780
Int. following stress	-0.830(0.316)	< 0.01	**	0.235	0.436	0.809
Int. following strong	4.990(0.397)	< 0.001	***	67.474	146.959	320.077
Other consonants	0.853(0.301)	< 0.01	**	1.302	2.347	4.232

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -3546. McFadden R^2 : 0.16056. Likelihood ratio test: $\chi^2 = 1356.5$, $p < 0.001$.

Most phonetic variants from /g/ occurred intervocalically (see Table 4.12). However, the proportion of elided variants was considerably smaller than for /b/ and /d/. In order to explore whether phonetic variant could be predicted from phonetic context, a MLR analysis was conducted. The phonetic contexts “following pause”, “following lateral”, “following nasal”, “preceding consonant” and “following other consonant” were collapsed into “around consonants”. The phonetic context “other” was not included in the analysis. The analysis was conducted with phonetic variant as the dependent variable (reference level: open approximant) and phonetic context as the independent predictor variable, in this case with four levels (reference level: preceding stress). The results for this analysis are summarized in Table 4.13.

The results from the MLR analysis showed that phonetic context is not a particularly good predictor of phonetic variant for /g/, since only around half of the phonetic contexts were able to predict whether the variant would surface as different from the reference level “open approximant”. For those contexts with significant results, the estimates and odds ratios do not necessarily align with the theoretical expectation of greater probability of finding elided variants in weaker contexts.

Table 4.12. Cross-tabulation of phonetic context against phonetic variant for variants from /g/ (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “PF”: plosives, fricatives and others).

Phonetic contexts	**EL**	**VA**	**OA**	**CA**	**PF**	**Sum**	**(%)**
Intervocalic preceding stress	48	161	310	53	45	617	(30.2%)
Intervocalic following stress	23	195	191	36	19	464	(22.7%)
Intervocalic	46	167	170	36	22	441	(21.6%)
Following lateral	0	6	33	20	11	70	(3.4%)
Following nasal	13	2	8	4	39	66	(3.2%)
Preceding consonant	22	15	3	0	2	42	(2.1%)
Following other consonant	0	1	3	0	1	5	(0.2%)
Following pause	0	0	1	0	1	2	(0.1%)
Other	42	51	137	47	59	336	(16.4%)
Sum	194	598	856	196	199	2043	

Table 4.13. Summary of results for an MLR analysis for /g/. Phonetic variant was defined as dependent variable and phonetic context as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	-1.865(0.155)	< 0.001	***			
Intervocalic	0.558(0.227)	< 0.05	*	1.119	1.748	2.729
Int. following stress	-0.251(0.270)	= 0.351		0.458	0.778	1.320
Around consonants	1.550(0.271)	< 0.001	***	2.768	4.709	8.010
Vocalic app. vs. Open app.						
(Intercept)	-0.655(0.097)	< 0.001	***			
Intervocalic	0.637(0.146)	< 0.001	***	1.421	1.891	2.518
Int. following stress	0.676(0.141)	< 0.001	***	1.492	1.966	2.590
Around consonants	-0.038(0.268)	= 0.887		0.569	0.963	1.629
Closed app. vs. Open app.						
(Intercept)	-1.766(0.149)	< 0.001	***			
Intervocalic	0.214(0.236)	= 0.365		0.780	1.239	1.968
Int. following stress	0.098(0.235)	= 0.678		0.696	1.102	1.747
Around consonants	1.073(0.291)	< 0.001	***	1.654	2.925	5.172
Other vs. Open app.						
(Intercept)	-1.930(0.160)	< 0.001	***			
Intervocalic	-0.115(0.277)	= 0.679		0.518	0.892	1.535
Int. following stress	-0.378(0.289)	= 0.190		0.389	0.685	1.207
Around consonants	2.048(0.255)	< 0.001	***	4.706	7.750	12.764

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -2232.5. McFadden R^2: 0. 039902. Likelihood ratio test: $\chi^2 = 185.57$, $p < 0.001$.

4.3.2. Word status

As displayed in Table 4.14 for /b/, vocalic and open approximants are more frequent in nonsense words, with most occurring as open approximants. In the case of words, elided, vocalic and open approximants make up a very similar proportion of variants. In order to explore the association between word status and phonetic variant for /b/, a chi-squared test was conducted. A significant statistical association between word status and phonetic variant was found (χ^2 = 115.56, df = 4, p < 0.001; Cramer's V = 0.205), with Cramer's V indicating a moderate effect size (Kotrlik, Williams & Jabor, 2011).

Table 4.14. Cross-tabulation of phonetic variant and word status for /b/. Within row percentages are shown below each frequency in brackets (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “P/F”: plosive or fricative).

Status	**EL**	**VA**	**OA**	**CA**	**P/F**	**Row Total**	**(%)**
Nonsense word	3 (1%)	103 (33%)	146 (46%)	24 (8%)	39 (12%)	315	(11%)
Word	651 (27%)	688 (28%)	679 (28%)	117 (5%)	290 (12%)	2425	(89%)
Column Total	654	791	825	141	329	2740	

To further explore this association, a MLR analysis was conducted (elided and vocalic approximants were collapsed into “highly lenited”), with phonetic variant as the dependent variable with four levels (reference level: open approximant) and word status as the independent predictor variable with two levels (reference level: words). The results showed that it was less likely for a nonsense word to contain a highly lenited variant or a fricative or plosive variant (see Table 4.15). These results are in line with predictions of less lenition for nonsense words in a word list task and more vocalic and open approximants in intervocalic contexts instead of fricatives or plosives.

The data for /d/ showed that vocalic and open approximants predominate for nonsense words, while elided variants are favoured for words, followed by vocalic approximants (see Table 4.16). Overall, lenition was considerably stronger in words than in nonsense words, despite the fact that all nonsense words had an intervocalic context. A chi-squared analysis was conducted to explore the association between word status and phonetic variant for /d/, which turned out to be significant (χ^2 = 217.92, df = 4, p < 0.001; Cramer's V = 0.248), with Cramer's V indicating a moderate effect size (Kotrlik, Williams & Jabor, 2011).

Table 4.15. Summary of results for MLR analysis for /b/, with phonetic variant as a dependent variable and word status as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Highly lenited vs. Open app.						
(Intercept)	0.679(0.047)	< 0.001	***			
Nonsense word	-0.999(0.136)	< 0.001	***	0.282	0.368	0.481
Closed app. vs. Open app.						
(Intercept)	-1.758(0.100)	< 0.001	***			
Nonsense word	-0.047(0.242)	= 0.846		0.594	0.954	1.533
Other vs. Open app.						
(Intercept)	-0.851(0.070)	< 0.001	***			
Nonsense word	-0.469(0.193)	< 0.01	*	0.428	0.625	0.914

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -3001.1. McFadden R^2: 0.0097164. Likelihood ratio test: $\chi^2 = 58.892$, $p < 0.001$.

Table 4.16. Cross-tabulation of phonetic variant and word status for /d/. Within row percentages are shown below each frequency in brackets (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “P/F”: plosive or fricative).

Status	**EL**	**VA**	**OA**	**CA**	**P/F**	**Row Total**	**(%)**
Nonsense word	37 (12%)	121 (38%)	118 (37%)	19 (6%)	25 (8%)	320	(9%)
Word	1408 (44%)	674 (21%)	462 (14%)	147 (5%)	545 (17%)	3236	(91%)
Column Total	1445	795	580	166	570	3556	

A MLR analysis with phonetic variant as the dependent predicted variable (reference level: open approximant) and word status as an independent variable (reference level: words) showed that it was less likely for /d/ to surface as any of these variants instead of open approximant. In other words, it is more likely for nonsense words to display open approximant variants (see Table 4.17).

Table 4.17. Summary of results for MLR analysis for /d/, with phonetic variant as a dependent variable and word status as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	1.114(0.054)	< 0.001	***			
Nonsense word	-2.274(0.196)	< 0.001	***	0.070	0.103	0.151
Vocalic app. vs. Open app.						
(Intercept)	0.378(0.060)	< 0.001	***			
Nonsense word	-0.353(0.143)	< 0.05	*	0.531	0.703	0.930
Closed app. vs. Open app.						
(Intercept)	-1.145(0.095)	< 0.001	***			
Nonsense word	-0.681(0.265)	< 0.05	*	0.301	0.506	0.850
Other vs. Open app.						
(Intercept)	0.165(0.063)	< 0.01	**			
Nonsense word	-1.717(0.229)	< 0.001	***	0.115	0.180	0.281

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -4986.3. McFadden R^2: 0.021562. Likelihood ratio test: $\chi^2 = 219.77$, $p < 0.001$.

In the case of /g/, open and then vocalic approximants were more frequent for both nonsense words and real words, showing a difference with respect to /b/ and /d/ where lenited variants were more frequent for words (see Table 4.18). For nonsense words, this was followed by closed approximants and for words, by elided variants. The results of a chi-squared test showed a significant statistical association between word status and phonetic variant ($\chi^2 = 46.892$, df = 4, $p < 0.001$; Cramer's V = 0.152), although Cramer's

V indicated a weak association (Kotrlik, Williams & Jabor, 2011).

Table 4.18. Cross-tabulation of phonetic variant and word status for /g/. Within row percentages are shown below each frequency in brackets ("EL": elided, "VA": vocalic approximant, "OA": open approximant, "CA": closed approximant, "P/F": plosive or fricative).

Status	**EL**	**VA**	**OA**	**CA**	**P/F**	**Row Total**	**(%)**
Nonsense word	8 (2%)	118 (30%)	167 (43%)	61 (16%)	36 (9%)	390	(19%)
Word	186 (11%)	480 (29%)	689 (42%)	135 (8%)	163 (10%)	1653	(81%)
Column Total	194	598	856	196	199	2043	

Table 4.19. Summary of results for MLR analysis for /g/, with phonetic variant as a dependent variable and word status as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Highly lenited vs. Open app.						
(Intercept)	-0.034(0.054)	= 0.532				
Nonsense word	-0.248(0.130)	< 0.1	.	0.605	0.78	1.007
Closed app. vs. Open app.						
(Intercept)	-1.630(0.094)	< 0.001	***			
Nonsense word	0.623(0.177)	< 0.001	***	1.318	1.86	2.636
Other vs. Open app.						
(Intercept)	-1.441(0.087)	< 0.001	***			
Nonsense word	-0.093(0.203)	= 0.647		0.612	0.91	1.357

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -2407.1. McFadden R^2: 0.0045109. Likelihood ratio test: $\chi^2 = 21.815$, $p < 0.001$.

A MLR analysis was conducted with relevelling (elided variants and vocalic approximants were collapsed into "highly lenited"). Phonetic variant was defined as the dependent variable (reference level: open approximant) and word status as the independent variable (reference level: words). The only significant result showed that it was more likely for /g/ to surface as closed rather than open approximant when in a nonsense word, which reinforces the idea that there is less overall lenition for /g/ (see Table 4.19).

4.3.3. Elicitation procedure

In the case of /b/, the proportion of lenited variants increased from word lists to texts, and from texts to the conversation (see Table 4.20). Word lists displayed the largest percentages of vocalic and open approximants, while elision predominated in texts and then even more in the semi-guided conversation. The higher percentage of plosive and fricative variants in the semi-guided conversation can be explained by the fact that phonetic contexts were not controlled in this task. A chi-squared test showed a significant statistical association between task and phonetic variant ($\chi^2 = 345.24$, df = 8, $p < 0.001$; Cramer's V = 0.251), with Cramer's V indicating a moderate effect size (Kotrlik, Williams & Jabor, 2011).

Table 4.20. Crosstabulation of phonetic variant and task for /b/. Within row percentages are shown below each frequency in brackets (“EL”: elided, “VA”: vocalic approximant, “OA”: open approximant, “CA”: closed approximant, “P/F”: plosive or fricative).

Task	EL	VA	OA	CA	P/F	Row Total	(%)
Word lists	52 (6%)	328 (36%)	395 (44%)	50 (6%)	75 (8%)	900	(33%)
Texts	219 (31%)	202 (28%)	199 (28%)	32 (4%)	66 (9%)	718	(26%)
Conversation	383 (34%)	261 (23%)	231 (21%)	59 (5%)	188 (17%)	1122	(41%)
Column Total	654	791	825	141	329	2740	

A MLR analysis was conducted to further explore the association between phonetic variant and task. Phonetic variant was defined as the dependent variable (reference level: open approximant) and task as an independent predictor variable with three levels (reference level: short texts). The results showed that, overall, task is not a particularly good predictor of phonetic variant, since only half of the comparisons were statistically significant (see Table 4.21). These results showed that it is less likely for /b/ to surface as elided, plosive or fricative instead of as an open approximant in word-lists, whereas the opposite is true for the semi-guided conversation.

Table 4.21. Summary of results for MLR analysis for /b/, with phonetic variant as a dependent variable and task as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	0.096(0.098)	= 0.328				
Word-lists	-2.123(0.177)	< 0.001	***	0.085	0.120	0.169
Conversations	0.410(0.129)	< 0.01	**	1.171	1.507	1.938
Vocalic app. vs. Open app.						
(Intercept)	0.015(0.100)	= 0.881				
Word-lists	-0.201(0.125)	= 0.107		0.641	0.818	1.045
Conversations	0.107(0.135)	= 0.426		0.855	1.113	1.449
Closed app. vs. Open app.						
(Intercept)	-1.828(0.190)	< 0.001	***			
Word-lists	-0.239(0.243)	= 0.324		0.489	0.787	1.266
Conversations	0.463(0.240)	< 0.1	.	0.993	1.588	2.542
Other vs. Open app.						
(Intercept)	-1.104(0.142)	< 0.001	***			
Word-lists	-0.558(0.190)	< 0.01	**	0.395	0.572	0.831
Conversations	0.898(0.173)	< 0.001	***	1.749	2.454	3.442

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -3831.1. McFadden R^2: 0.048317. Likelihood ratio test: $\chi^2 = 389.02$, $p < 0.001$.

The results for variants of /d/ replicate those observed for /b/: vocalic and open approximants were more frequent in word-lists, while elided variants were more frequent both in texts and the semi-guided conversation (see Table 4.22). The rise of plosive and fricative variants in the last two tasks can be explained by the fact that word lists only contained intervocalic instances of /d/, whereas in the other tasks more phonetic contexts were present. A chi-squared analysis showed a significant statistical association between phonetic variant and task (χ^2 = 622.68, df = 8, p < 0.001; Cramer's V = 0.296), with Cramer's V indicating a moderate effect size (Kotrlik, Williams & Jabor, 2011).

Table 4.22. Crosstabulation of phonetic variant and task for /d/. Within row percentages are shown below each frequency in brackets ("EL": elided, "VA": vocalic approximant, "OA": open approximant, "CA": closed approximant, "P/F": plosive or fricative).

Task	EL	VA	OA	CA	P/F	Row Total	(%)
Word lists	156 (17%)	389 (42%)	280 (30%)	33 (4%)	63 (7%)	921	(26%)
Texts	447 (50%)	145 (16%)	100 (11%)	53 (6%)	144 (16%)	889	(25%)
Conversation	842 (48%)	261 (15%)	200 (11%)	80 (5%)	363 (21%)	1746	(49%)
Column Total	1445	795	580	166	570	3556	

A MLR analysis was conducted to further explore the association between phonetic variant and task. The analysis was set with phonetic variant as the dependent variable with five levels (reference level: open approximant) and task as the independent predictor variable with three levels (reference level: short texts). The results for this analysis are summarized in Table 4.23. Overall, task turned out to be a relatively poor predictor of phonetic variant for /d/, since most comparisons were not statistically significant. When it did – for elided, closed approximants, and fricatives and plosives –, the results confirmed that open approximants are most likely to occur in word lists.

The results for variants from /g/ showed a different pattern from those for /b/ and /d/. In the case of word lists, open and vocalic approximants are again more frequent, but elided variants did not overtake the other categories in texts and the semi-guided conversation; instead, open approximants remain the most frequent variant, and there is a gradual increase in the proportion of lenited variants (see Table 4.24). Once more, higher degrees of formality and attention to speech can account for these differences (references). A chi-squared analysis revealed a significant statistical association between task and phonetic variant ($\chi^2 = 189.04$, df = 8, $p < 0.001$; Cramer's V = 0.215), with Cramer's V indicating a moderate effect size (Kotrlik, Williams & Jabor, 2011).

Table 4.23. Summary of results for MLR analysis for /d/, with phonetic variant as a dependent variable and task as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	1.497(0.111)	< 0.001	***			
Word-lists	-2.082(0.149)	< 0.001	***	0.093	0.125	0.167
Conversations	-0.060(0.136)	= 0.659		0.722	0.942	1.229
Vocalic app. vs. Open app.						
(Intercept)	0.372(0.130)	< 0.01	**			
Word-lists	-0.043(0.152)	= 0.778		0.712	0.958	1.290
Conversations	-0.105(0.160)	= 0.511		0.657	0.900	1.232
Closed app. vs. Open app.						
(Intercept)	-0.635(0.170)	< 0.001	***			
Word-lists	-1.503(0.250)	< 0.001	***	0.136	0.222	0.363
Conversations	-0.281(0.215)	= 0.191		0.495	0.755	1.151
Other vs. Open app.						
(Intercept)	0.365(0.130)	< 0.01	**			
Word-lists	-1.856(0.191)	< 0.001	***	0.108	0.156	0.227
Conversations	0.231(0.157)	= 0.141		0.926	1.260	1.715

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -4785.1. McFadden R^2: 0.061042. Likelihood ratio test: $\chi^2 = 622.17$, $p < 0.001$.

Table 4.24. Crosstabulation of phonetic variant and task for /g/. Within row percentages are shown below each frequency in brackets ("EL": elided, "VA": vocalic approximant, "OA": open approximant, "CA": closed approximant, "P/F": plosive or fricative).

Task	EL	VA	OA	CA	P/F	Row Total	(%)
Word lists	17 (2%)	291 (32%)	406 (45%)	103 (11%)	80 (9%)	897	(44%)
Texts	50 (9%)	172 (29%)	263 (45%)	50 (9%)	53 (9%)	588	(29%)
Conversation	127 (23%)	135 (24%)	187 (34%)	43 (8%)	66 (12%)	558	(27%)
Column Total	194	598	856	196	199	2043	

A MLR analysis was prepared with phonetic variant as the dependent variable with five levels (reference level: open approximant) and task as the independent predictor variable with three levels (reference level: short texts). The results for this analysis are summarized in Table 4.25. Overall, task was a poor predictor of phonetic variant, with the majority of comparisons failing to reach significant values. Those that did showed that elided variants were less likely than open approximants to appear in word-lists and more likely to surface in semi-guided conversations; also, they showed that plosive and fricative variants were more likely to appear in the semi-guided conversation.

Table 4.25. Summary of results for MLR analysis for /g/, with phonetic variant as a dependent variable and task as predictor. Estimates, their statistical significance, standard errors (SE), odds ratios and confidence intervals for the odds ratios are provided.

Comparisons	**Estimate (SE)**	***p***		**Lower CI (2.5%)**	**Odds ratio**	**Upper CI (97.5%)**
Elided vs. Open app.						
(Intercept)	-1.660(0.154)	< 0.001	***			
Word-lists	-1.513(0.292)	< 0.001	***	0.124	0.220	0.390
Conversations	1.273(0.192)	< 0.001	***	2.450	3.572	5.209
Vocalic app. vs. Open app.						
(Intercept)	-0.425(0.098)	< 0.001	***			
Word-lists	0.092(0.125)	= 0.462		0.859	1.096	1.399
Conversations	0.099(0.150)	= 0.509		0.823	1.104	1.480
Closed app. vs. Open app.						
(Intercept)	-1.660(0.154)	< 0.001	***			
Word-lists	0.289(0.190)	= 0.128		0.920	1.334	1.935
Conversations	0.190(0.229)	= 0.406		0.772	1.210	1.894
Other vs. Open app.						
(Intercept)	-1.602(0.151)	< 0.001	***			
Word-lists	-0.022(0.194)	= 0.908		0.669	0.978	1.430
Conversations	0.560(0.208)	< 0.01	**	1.166	1.751	2.632

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1. Log-Likelihood = -2765.3. McFadden R^2: 0.032765. Likelihood ratio test: $\chi^2 = 187.34$, $p < 0.001$.

4.3.4. Internal word frequency[12]

A subset of instances of /b/ occurring in semi-guided conversations was isolated to explore the relationship between word frequency and phonetic variant. Treating word label as a categorical variable, frequencies were generated for each lexical item, along with the mean phonetic variant of those instances (assuming an underlying lenition continuum ranging from 1 for elided variants to 5 for plosives and fricatives). Internal lexical frequency was then sorted into descending order and the first 100 frequencies were selected.

12 Word frequency is "internal" in the sense that it was calculated for the present corpus, i.e., it was not extracted from existing lexical frequency lists.

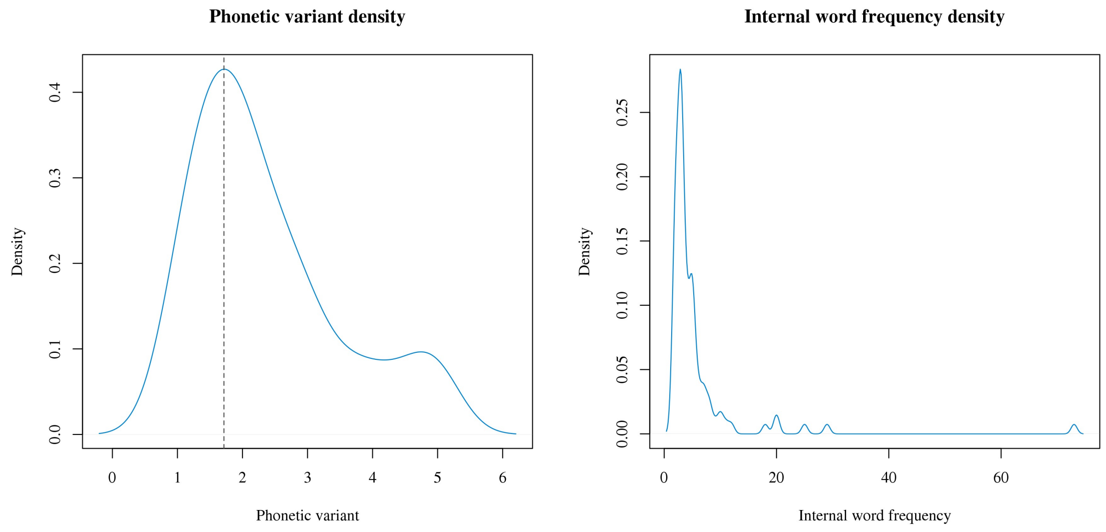

Figure 4.6. Kernel density plots for the 100 most frequent lexical items from /b/. Left-hand side panel: phonetic variant density. Right-hand side panel: internal word frequency density.

Kernel density plots showed that vocalic approximants and elided variants are more frequent amongst the 100 most frequent lexical items, with density peaking at a single mode at 1.71 (see left-hand panel from Figure 4.6). In the case of internal word frequency, a relatively small number of lexical items displayed frequencies different from 0 density, which makes it difficult to propose any general pattern of association between lexical frequency and phonetic variant (see right-hand panel from Figure 4.6).

Kernel density plots for data from the 100 most frequent lexical items for /d/, generated following the same methods described above, showed that elided variants were more frequent, with density showing the largest mode at 1.22 (see left-hand panel from Figure 4.7). A secondary mode and a plateau suggested that the preference for lenited and elided variants was not clear-cut. As was observed for /b/, only a few items displayed frequencies different from floor density (see right-hand panel from Figure 4.7).

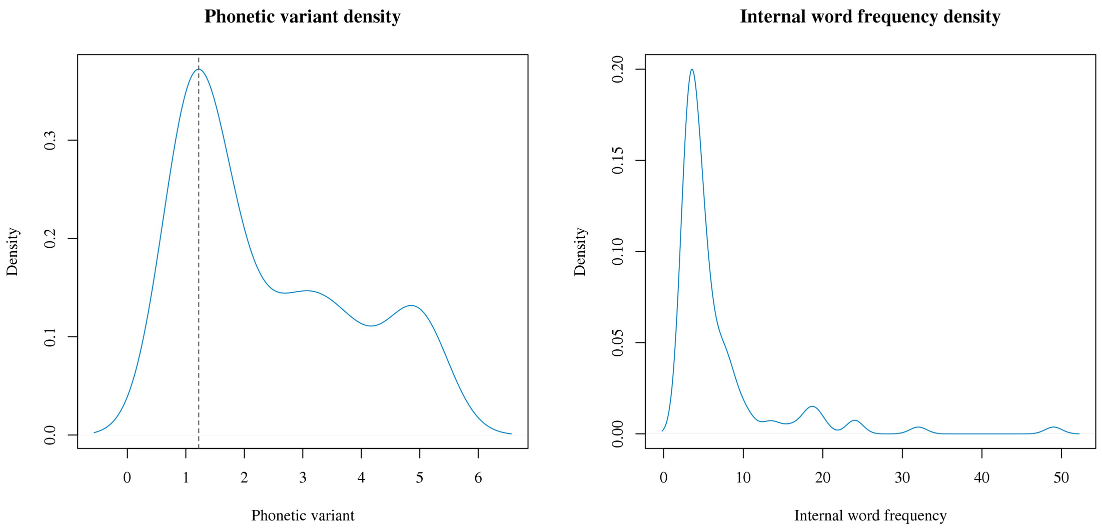

Figure 4.7. Kernel density plots for the 100 most frequent lexical items from /d/. Left-hand side panel: phonetic variant density. Right-hand side panel: internal word frequency density.

Kernel density plots were also prepared for a subset of the 100 most frequent lexical items containing /g/, in order to explore the relationship between internal lexical frequency and phonetic variant (see Figure 4.8). As shown by the vertical line, density peaks at 2.04, implying that vocalic approximants were more frequent. The fact that a single mode is observable suggests that there was a relationship between lexical frequency and phonetic variant. In this case, higher lexical frequency was associated with lenition, but to a lesser degree than for /b/ and even lesser than for /d/.

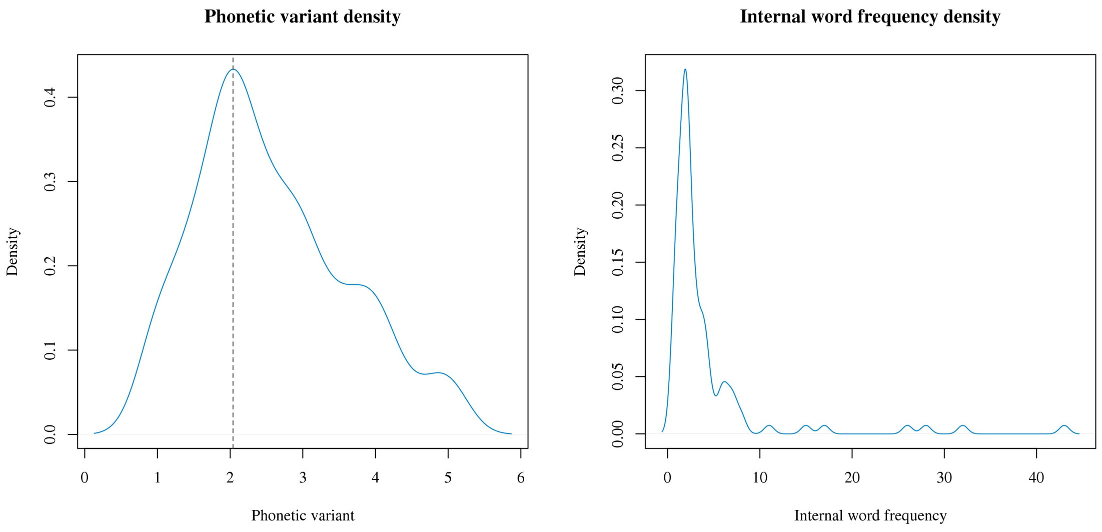

Figure 4.8. Kernel density plots for the 100 most frequent lexical items from /g/. Left-hand side panel: phonetic variant density. Right-hand side panel: internal word frequency density.

4.4. Summary of results

From a database of 8339 instances of /b d g/, which included three data elicitation tasks (word-lists, texts and semi-guided conversation), /d/ was the most frequent phoneme accounting for 42.6% of all instances, followed by /b/ (32.9%) and /g/ (24.5%). These results replicate previous findings for Chilean Spanish as well as from other dialects (Pérez, 2003). The aggregated results for /b d g/ also showed that open approximants, vocalic approximants and elided variants were by far the most frequent, all with percentages around 27%. If the allophones are examined by phoneme category, open and vocalic approximants are more frequent for /b/, although a high percentage of elision was found too (23.9%). In the case of /d/, elided variants were more frequent (40.6%), nearly doubling the next categories (vocalic and open approximants). Finally, open approximants were more frequent for /g/ (41.9%), followed by vocalic approximants. If /b d g/ are ranked by degree of lenition, /d/ displays the highest degree, followed by /d/ and then by /g/, in agreement with previous reports for Argentinian and Chilean Spanish (Colantoni & Marinescu, 2010; Pérez, 2007). In terms of phonetic context, aggregated for /b d g/, near to 70% of tokens were located in an intervocalic context; the next highest phonetic context was following nasals, with 5.4% of all instances.

4.4.1. Acoustic properties[13]

Duration

In line with previous findings, duration of the approximant variants from /b/ was relatively short (between 48 to 65 ms). Statistical results on the normalized data showed a main effect of phonetic variant on the normalized duration for /b/; also, significant differences were found between vocalic and open approximants, and between vocalic and closed approximants. No significant differences were found between open and closed approximants. In the case of /d/, absolute duration values were, on average, around 51 to 60 ms. Statistical analyses conducted on the normalized results showed a significant main effect of phonetic variant. Also, significant differences were found between vocalic and open approximants, between vocalic and closed approximants, but not between open and closed approximants. Variants of /g/ had slightly longer duration values when compared to the other categories (between 53 to 66 ms). A significant effect of phonetic category was found on the normalized duration values, as well as significant differences between vocalic and open approximants, but not between vocalic and closed approximants or open and closed approximants.

Intensity

Results were similar across /b d g/. Absolute intensity values ranged from 36 to 51 dB, with more open variants displaying higher intensity values than closer ones. Additionally, there was a significant main effect of phonetic variant and significant differences were found between all allophonic categories within each variant.

13 As explained elsewhere (see the introduction to “4.2. Acoustic properties of approximant variants of /b d g/”), elided, plosive and fricative variants were excluded from acoustic analyses.

F1

As expected, different variants had different ranges of formant frequencies. For /b/, vocalic approximants showed a mean non-normalized F1 value of 474 Hz, open approximants 418 Hz, and closed approximants 346 Hz. For /d/, vocalic approximants showed mean non-normalized values of 477 Hz, open approximants of 389 Hz and closed approximants of 384 Hz. For /g/, the non-normalized means showed a value of 423 Hz for vocalic approximants, 381 Hz for open approximants and 382 for closed approximants (mean values of 412 Hz, 375 Hz and 351 Hz, respectively). For each phoneme, there was a significant main effect of phonetic variant, as well as significant differences between vocalic and open approximants and vocalic and closed approximants.

F2

The non-normalized F2 results for /b/ showed mean values of 1471 Hz, 1455 Hz and 1462 Hz for vocalic, open and closed approximants. No main effect of phonetic variant on the Lobanov-normalized F2 values was found, nor were there significant differences between sub-categories. Mean non-normalized F2 values for /d/ were 1559 Hz for vocalic approximants, 1574 Hz for open approximants and 1639 Hz for closed approximants. In contrast to /b/, a main effect of phonetic variants on normalized F2 values for /d/ was found, along with significant differences between vocalic and closed approximants, and between open and closed approximants (but not between vocalic and open approximants). For /g/, the mean raw F2 values were 1592 Hz for vocalic approximants, 1439 Hz for open approximants and 1566 Hz for closed approximants. As for /d/, there was a significant main effect of phonetic variant on the normalized F2 values for /g/. Significant differences were also detected between vocalic and open approximants, but not for any other comparison.

4.4.2. Indexical variables

Phonetic context

Most instances of /b/ surfaced intervocalically (in total, 68.3%), followed by instances of /b/ after nasals (4.2%). A MLR analysis showed that, overall, phonetic context was a good predictor of phonetic variant for /b/. Moreover, estimates and odds ratios confirmed that lenited variants were more likely to surface in weaker phonetic contexts (e.g., intervocalically). In the case of /d/, in which elided and vocalic variants are considerably more frequent, most variants surfaced in an intervocalic context (66.8%), followed by a relatively high number of /d/ tokens following nasal consonants (7.5%). The MLR analysis showed that phonetic context was also a good predictor of phonetic variant for /d/, and more lenited variants surface in weaker phonetic contexts. As for /g/, again, most variants surfaced in intervocalic contexts (74.5%). In the case of /g/, the MLR analysis showed that phonetic context was not a particularly good predictor of phonetic variant, since several phonetic contexts were not able to predict how phonetic variant would surface. For those contexts where it did, not all estimates and odds ratios aligned with theoretical expectations of weaker variants in weaker contexts.

Word status

In the case of /b/, more elided variants were found in words than in nonsense words (the most frequent variant for nonsense words were open approximants). As well as a significant statistical association between word status and phonetic variant, an MLR analysis showed that it was less likely for a nonsense word to contain highly lenited variants. In the case of /d/, overall, lenition was considerably stronger in words than in nonsense words. A chi-squared analysis revealed a significant statistical association between word status and phonetic variant. As for /b/, an MLR analysis for the results from /d/ also showed that it was very unlikely for an elided variant to surface in nonsense words, and additionally that open approximants were more likely for this category. Vocalic approximants were more frequent for /g/ in both words and nonsense words, although elision was considerably higher in words. The results from a chi-squared test revealed a significant statistical association between word status and phonetic variant for /g/. Overall, word status was not a particularly good predictor of phonetic variant. The only significant result showed that it was more likely for /g/ to surface as closed approximant (instead of open) in nonsense words, reinforcing the idea of less overall lenition for /g/.

Elicitation task

Inspection of the results for /b/ showed that the proportion of lenited variants increased from word lists to texts, and then to the semi-guided conversation, where the highest proportion of elided variants and open approximants was found. A chi-squared analysis showed a significant statistical association between task and phonetic variant. An MLR analysis on the results showed that, overall, task was not a good predictor of phonetic variant. The results that reached significance showed that it was less likely for /b/ to surface as elided in word-lists, while the opposite was true in the semi-guided conversation. In the case of /d/, more lenited variants were also more frequent in text and semi-guided conversation. As well as a significant statistical association between phonetic variant and task, shown by a chi-squared analysis, and MLR analysis showed that task was a relatively poor predictor of phonetic variant for instances for /d/. In those few comparisons where statistically significant results were found, it was confirmed that open approximants were more likely in word-lists (as opposed to texts and the semi-guided conversation). The results for /g/ showed a different pattern: open approximants were more frequent across tasks, although elided variants and open approximants increased significantly in the semi-guided conversation. A chi-squared analysis showed a significant statistical association between task and phonetic variant from /g/. An MLR analysis showed that task was a poor predictor of phonetic variant. Those few comparisons in which significant results were found showed that it was less likely for elided variants to surface in word-lists, and more likely to surface in semi-guided conversations. Overall, the analyses showed that although there were some differences in the surface variants for each phoneme, task was a poor predictor of phonetic variant for /b d g/.

Internal word frequency

Graphical visualizations for the 100 most frequent lexical items from the corpus showed that vocalic and elided variants were more frequent for high-frequency items from /b/, elided variants for high-frequency items from /d/, and vocalic variants for high-frequency items from /g/. Overall, more frequent variants were associated to more lenited variants (elided or vocalic), although this trend was less strong for /g/. The fact that for all three consonants a relatively small number of lexical items displayed frequencies different from near 0 makes it difficult to confirm a clear pattern of association between lexical frequency and phonetic variant.

4.5. Discussion

4.5.1. Explaining the variation of Chilean Spanish /b d g/

Chilean Spanish: a particularly lenited dialect

The acoustic results summarized above show that Chilean Spanish can be characterized as a dialect with a particularly high degree of lenited and elided variants. Although this trend is true for all three phonological categories, it finds its extreme in /d/, for which elision was the most frequent variant, followed by /b/, in which highly lenited variants predominated, and then by /g/, in which open approximants were more frequent. As mentioned already, these results agree with previous reports made for Chilean Spanish (Cepeda, 1991; Pérez, 2007), and place Chilean Spanish in one of the extremes in a scale ranging from those dialects affected by fortition –e.g., Honduran and Costa Rican (Amastae, 1989; Carrasco, Hualde & Simonet, 2012)– and those affected by lenition –e.g., Miami Spanish (Hammond, 1976). All things considered, it can be hypothesized that Chilean Spanish is likely being affected by lenition pressures, and thus a lenition account or one including underspecified phonological units would probably better explain the current data (cf. Harris, 1969; Piñeros, 2002; Lozano, 1978; Mascaró, 1984), although only diachronic studies would be able to settle this.

Interpreting acoustic results in the light of research precedents

In general, the acoustic results are in line with previous research and theoretical predictions. In the case of duration, besides the fact that approximant consonants are generally short, open variants were shorter than closed approximants, supporting the assumption that constriction degree is positively correlated to duration (Romero Gallego, 1995). As to intensity, the acoustic results also confirm that more constricted variants display lower relative intensity values when compared to more open variants, which has been attested numerous times in the specialized literature (e.g, Hualde, Shosted & Scarpace, 2011; Simonet, Hualde & Nadeu, 2012). The results for F1 showed that, overall, more open variants display higher F1 values. This is clearly the case for /b/, and also for /d/ and /g/ except in the comparison between open and closed approximants. This trend might be explained by a hypothetical positive correlation between the distance of the articulators in the oral cavity and F1 values, such as the one observed for vowels (Ladefoged, 2003), but also more generally for differences observed between vowels and consonants, whereby more narrow constrictions result in lower F1 values (Stevens, 2002). No clear patterns of F2 differences were observed between approximant variants of /b d g/: while in /d/ more open realizations display lower F2 values, both in /b/ and /g/ vocalic approximants present the highest values, followed by closed approximants and then by open approximants.

Are subcategories of approximant consonants warranted?

In an attempt to systematize the variation observed in approximant consonants, and also to distinguish unequivocally between approximants and fricatives, three sub-categories have been proposed for spirant approximant consonants: vocalic, open and closed (Martínez-Celdrán, 2004, 2013; Martínez-Celdrán & Regueira, 2008; see “2.5.1. Defining the term 'approximant'” and “3.3.4. Segmentation, labelling and coding”). So far in this dissertation I have employed these categories uncritically as a means to compartmentalize a hypothetical continuum of realizations into three stages of lenition, but it remains to be determined whether they are substantiated by acoustic and statistical evidence. However, the evidence here supports this position. Evidence for these three stages of lenition comes from main effects of phonetic variant on duration, intensity and F1, for the three phonological categories. Moreover, *post hoc* analyses found statistically significant differences between the allophonic categories for all comparisons regarding intensity, and for most comparisons regarding duration and F1 (in these, vocalic approximants are significantly different from open and closed approximants, but these last two variants are not always statistically different). The fact that no main effect of phonetic variant on F2 values for /b/ was found, that no clear trend was observed in the acoustic results for F2, along with the higher degree of overlap between the allophonic categories for this variable indicates that this acoustic variable is not relevant for the description of allophonic variation in approximant consonants from /b d g/. Taken together, these results, along with those presented in previous sub-sections, suggest that the variability relevant to identifying the approximant sub-categories on a scale from more constriction to lenition is concentrated in the acoustic variables of duration, intensity and F1.

Independent variables affecting /b d g/ variation

The results found for phonetic context replicate the main trends observed for other variants from Spanish (e.g., Carrasco, Hualde & Simonet, 2012; Simonet, Hualde & Nadeu, 2012). Generally speaking, prominent phonetic contexts (see Escure, 1977) and domain-medial prosodic contexts (see Fougeron & Keating, 1997; Keating, Cho, Fougeron & Hsu, 2004; Cho & Keating, 2009) favoured more constricted variants, and made lenition less likely in Chilean Spanish.

In the case of word status, words favoured more lenited variants, which is in line with the expectations given higher attention to speech in that particular task (Labov, 1972). In terms of the type of elicitation task used, results showed that lenition was more likely to happen in the semi-guided conversation, as opposed to the word list and texts, again in agreement with previous accounts (Carrasco, Hualde & Simonet, 2012; Johnson, 2004). These results can again be explained by higher degrees of formality and attention to speech in the word lists task (Labov, 1972). Finally, regarding word frequency, the results confirm previous reports for Spanish and other languages finding that word-frequency is positively correlated to degree of lenition (Bybee, 2002, 2003; Eddington, 2011; Brown, 2013).

4.5.2. On methodological standards

Segmentation

In relation to segmentation, a consistent and explicit set of criteria was followed throughout the data collection stage, which is particular important given that approximants are inherently difficult to segment (see Turk, Nakai & Sugahara, 2006). In this, this approach follows previous work that has developed similar protocols using information from the waveform and spectrogram to reach a sensible segmentation hypothesis (Kingston, 2008). However, because no automated segmentation method such as forced aligners or cross researcher segmentation agreement were attempted, it is not possible to evaluate the efficacy of our manual segmentation approach.

Normalization procedures

Only one normalization procedure for duration was implemented, mainly due to the annotation protocol, which did not include a segmental or syllabic level segmentation of the whole signal, but instead only the target segment and its neighbours. The normalization procedure involved calculating the relative duration of the target segment with respect to that of the combined duration of the segment plus its two neighbours (Martínez Celdrán, 2013). Normalizing by speech rate was thus not an option. Similarly, normalizing by comparing the duration of the target segment to that of the host word did not seem like a better alternative, since in this corpus words had very different syllabic lengths. As explained in “3.4.1. Duration”, normalizing duration did not seem to affect the source of variation encoding constriction degree differences, while it reduced skew and kurtosis from the distributions, making them closer to normal. Overall, the normalization of duration was effective in reducing undesired sources of variation, and preserving those of interest.

In the case of intensity, relative intensity measurements are the preferred method to normalize absolute values, although a note of caution from using such methods has been raised (Hualde, Shosted & Scarpace, 2011). Five methods were implemented as described in the literature –intensity ratio, perseveratory intensity difference, anticipatory intensity difference, maximum velocity and minimum velocity (see “3.4.2.

Intensity” for a full description of the methods). Their results were then evaluated via graphic visualizations (at which stage minimum velocity was discarded) and LDA and QDA analyses predicting phonetic variant. The analyses revealed that all evaluated relative intensity indexes performed better than baseline intensity, and thus were successful at maximizing the desired sources of variation, and perhaps at removing undesired ones. Amongst the methods, intensity ratio performed marginally better, and thus the results obtained from this normalization method were used for subsequent statistical analyses in our data.

To my knowledge, this is the first study in which oral formant normalization procedures have been attempted on spirant approximants. Of course, oral formant normalization procedures for vowels are well known (for a review, see Adank, Smits & van Hout, 2004), but it was unclear whether they would also be effective in removing undesired sources of variation from acoustic data obtained from spirant approximant consonants. Four methods were implemented –Lobanov's Z-score transformation, Nearey 1, Nearey 2, and Labov's modification of Nearey's (see “3.4.3. Formants” for details). Again, visualizations of the results and LDA and QDA analyses predicting sex and phoneme categories were conducted on the normalized data to observe whether the methods were successful at removing the variation originating from sex differences, and at preserving or maximizing the variation originating from phonemic differences. In short, this was the case for all normalization methods. Although all of them performed similarly, Lobanov was selected given a slight advantage as revealed by the QDA sex classification, confirming previous results for vowel normalization (Adank, Smits & van Hout, 2004).

I would like to finalize this sub-section with a call for methodological caution. It is not unusual to find in the specialized literature that results like the ones above are taken blindly as rules-of-thumb to be applied to new corpora of similar data. However, as these results demonstrate, with a few exceptions, the methods performed almost equally well, which indicates that comparing them and evaluating their behaviour for each specific set of data is a safer approach. Consequently, although intensity ratio and Lobanov were the methods that showed the best results at removing undesired sources of variation and at retaining desired ones, it is not necessarily the case that they should be adopted as gold standards.

4.5.3. Implications for speech perception

All subsequent chapters deal in one way or another with how Chilean native listeners perceive approximant variants from /b d g/. It is expected that the perception of these consonants is mediated by the expectations that listeners have regarding what constitutes normal realizations or, if not, plausible realizations of approximants. In the following subsections the relevance of the production results for perception studies will be discussed briefly.

Differences between /b d g/

The results summarized above showed that the phonetic variants from /b/, /d/ and /g/ are not equally distributed. For /b/, open approximants are more frequent (30.1%), followed by vocalic approximants (28.9%) and elided variants (23.9%). For /d/, elided variants display the largest percentage by far (40.6%), followed by vocalic approximants (22.4%) and open approximants (16.3%). Finally, for /g/, most variants surfaced as open approximants (41.9%), followed by vocalic approximants (29.3%), with all other variants, included elided variants, having similar percentages (around 9.6%). As can be seen, /d/ displayed the highest degree of lenition and elision, followed by /b/, with fewer but still a considerable number of elided variants, and finally by /g/, the least lenited category, in which canonical approximants predominated, followed by vocalic approximants.

It is highly likely that these differences will have an impact on perception. Assuming that listeners are attuned to these trends, it can be hypothesized that they will tolerate more lenition for /d/ than for /g/. If this is true, then listeners might rely less on acoustic evidence to perceive approximant variants from /d/, when compared to /b/ or /g/; conversely, listeners may rely more on alternative sources of evidence to perceive approximants from /d/, and less so for /b/ and /g/, which are better backed-up by acoustic evidence.

Another difference between /b d g/ that might have an impact on perception is that /b/ displays both bilabial and labiodental realizations (e.g., Sadowsky, 2010). No attempts to distinguish between these two places of articulation were made in the production study, under the assumption that listeners are not able to perceive the difference for approximant realizations[14]. If listeners are able to discriminate between [ʋ] and [β̞], then the design of perception experiments for the perception of /b/ should take this variation into account. This is investigated in Chapter 5.

Investigating the perception of variants of /b d g/

The next chapter presents experiments investigating the perception of /b d g/ variants by Chilean Spanish listeners, using carefully constructed acoustic continua. Although the use of a continua from spirant approximant to elision contrasts with those used in most previous perception studies, this chapter has shown that such variation exists naturally in the speaker's realizations and thus likely also in the listener's expectations. Given that phonetic context, word status and lexical frequency have been shown to affect perception, they will have to be carefully controlled in order to provide listeners with an environment in which the full range of variation is plausible, and to control the potential effect of some confounding variables in the results.

14 I refer to [ʋ] versus [β̞]. Listeners probably do discriminate between [v] and [b], but this is of no relevance for this thesis.

Chapter 5
Perception of the bilabial–labiodental contrast within /b/

5.1. Introduction

Variants of /b/ in Spanish and Chilean Spanish

As described in section “2.5.2. Spanish /b d g/ and their approximant variants”, /b/ displays several variants ranging from voiced plosives to elision, and this variation differs between dialects. Approximant realizations of /b/ are relatively short, with average durations between 30 and 60 ms (Almeida & Pérez Vidal, 1991; Martínez Celdrán, 1984; 2013), and their intensity is inversely proportional to their degree of constriction. Formant values, display F1 values around 405 Hz, and around 1080 Hz for F2 (Almeida & Pérez Vidal, 1991). In the case of Chilean Spanish (see section “2.5.3. Chilean Spanish /b d g/”), several variants have been reported, including plosives, fricatives, relaxed fricatives, approximants and elided realizations. More constricted variants surface in some phonetic contexts such as word-initial or after nasals, and highly lenited and elided variants in other contexts, such as intervocalically (e.g., Cepeda, 1991).

Earlier reports dismiss the existence of labiodental variants of /b/ for any dialect of Spanish (see Sadowsky, 2010). However, more recent studies using video recordings of subjects' lips have demonstrated that labiodental variants of /b/ not only exist in Chilean Spanish (Cepeda, 2001; Borland Delorme, 2004), but often are more frequent than bilabial realizations, with whom they coexist in free variation, at least for the most part (Sadowsky, 2010; Vergara Fernández, 2011, 2013; Vergara & Pérez, 2013). These studies have also shown that speakers do not tend to instantiate orthographic “b” as bilabial variants or “v” as labiodental variants (“b” for [b] and [β̞], and “v” for [ʋ]), both in literate and preliterate individuals (Sadowsky, 2010; Vergara Fernández, 2011; 2013; Vergara & Pérez, 2013). No precedents exist for the perception of approximant variants of /b/ in Spanish.

Summary of results from production study

The results from the production study (see Chapter 4) showed that most realizations of /b/ are open or vocalic approximants, or elided variants (82.9%). No attempts were made in that study to distinguish bilabial from labiodental realizations. On average, approximant variants showed duration values between 48 and 65 ms, and the normalized duration differences between phonetic variants were statistically significant, except in the case of the comparison between open and closed approximants. Intensity was found to be inversely proportional to the degree of constriction, with less constricted variants showing higher intensity values. All normalized intensity differences between approximant variants were statistically significant. In the case of F1, vocalic approximants showed mean values of 473.9 Hz, open approximants of 417.7 Hz and closed approximants of 346.3 Hz. Again, all the differences between normalized F1 values of approximant variants were found to be statistically significant. No clear trends were observed for F2, and no significant differences were found between the normalized F2 values for approximant variants of /b/. The average F2 value for the three approximant variants was 1462.7 Hz, considerably higher than previous reports (Almeida & Pérez Vidal, 1991).

Most phonetic variants of /b/ surfaced in intervocalic contexts (68.3%), in which elided, vocalic and open approximants predominated over closed approximants, plosives and fricatives. An MLR analysis with phonetic variant as the dependent variable and phonetic context as an independent predictor showed that phonetic context reliably predicted realizations of /b/ and that it was more likely for lenited variants to surface in weaker phonetic contexts. Word status was also found to be a good predictor of phonetic variants of /b/ in an MLR analysis. In particular, highly lenited variants, and fricative and plosive variants were less likely to be found in nonsense words. A significant association between experimental task and phonetic variants of /b/ was also identified. However, an MLR analysis showed that task was not a good predictor of phonetic variant in our sample. Finally, a link between lexical frequency and lenition was found, with more lenited variants tending to have higher lexical frequencies.

Acoustic and perceptual differences between [β̞] and [ʋ]

To date, no information exists about potential acoustic differences between Spanish [β̞] and [ʋ]. Those studies that do report acoustic data for approximant variants of /b/ did not factor place of articulation in their analyses, nor attempted to exclude labiodental realizations from their samples (Almeida & Pérez Vidal, 1991; Martínez Celdrán, 1984; 2013, and Chapter 4 of our own study). In my results, no evidence of a bimodal distribution was observed in any of the raw and normalized acoustic variables explored for the approximant variants of /b/ in Chapter 4, which suggests that there is no *a priori* reason to believe that place of articulation is encoded acoustically, although it is not possible to rule out such an effect.

There is no information either regarding the perception of approximant variants of /b/, that is, whether listeners are able to identify [β̞] and [ʋ] or discriminate between them. If acoustic differences exist, and Chilean Spanish listeners are able to identify and discriminate [β̞] and [ʋ], this variability would have to be taken into account and controlled for in designing perception experiments involving approximant variants of /b/, as those to be discussed later in Chapter 6 and Chapter 7. More importantly, if a perceptual distinction between [β̞] and [ʋ] exists, it could be used in Chilean Spanish to encode linguistic and non-linguistic information, which would increase the chances of entering confounding variables into the results of perception experiments for /b/.

5.2. Aims

The following experiments aim to determine whether native Chilean Spanish listeners are able to identify and discriminate [β̞] and [ʋ]. If listeners display clear identification and discrimination patterns, this variation has to be taken into account in subsequent perception experiments, as it might encode linguistic or extra-linguistic information that could act as a confounding variable. If the opposite is true, i.e., null results are found for both identification and discrimination tasks, then it can be posited that listeners are not sensitive to this variation and thus it can be disregarded safely in subsequent perception experiments involving /b/.

5.3. Methods

5.3.1. Participants

Thirty one adult native monolingual Chilean Spanish speakers took part in this experiment (6 males, 25 females; mean age 20.6 years). Participants were all undergraduate university students residing in Santiago, Chile. None of the participants reported any hearing, speech, language or other impairment. Participants read an information sheet, filled in a questionnaire and signed a consent form prior to the experiment. Participants were compensated for their participation.

5.3.2. Stimuli

Several instances of the minimal pairs [ˈsa.β̞a] - [ˈsa.ʋa] and [ˈga.β̞a] - [ˈga.ʋa] were recorded by the author, a monolingual native Chilean Spanish speaker. The recordings were conducted in a sound-isolated booth using a Rode NT1A condenser microphone. The signal was sent via an RME Fireface UC interface to a desktop computer running RME TotalMix mixer. Recordings were set to a sampling frequency of 44100 Hz and 16 bit depth. Video recordings were taken simultaneously on a Canon LEGRIA HF G30. Eight stimuli were selected for each nonsense word pair to be used in a natural identification task. Inspection of the waveforms and spectrograms in *Praat* (Boersma & Weenink, 2015) ensured that the approximant consonants of the stimuli were indeed open approximants (Martínez-Celdrán & Regueira, 2008). Frame by frame analysis of the video recordings was used to confirm the place of articulation of each token. Mean intensity was scaled to 70 dB. Averaged acoustic values for the approximant consonants from each nonsense word are provided in Table 5.1 (intensity corresponds to the minimum intensity found within the consonant; fundamental frequency, oral formant values from F1 to F5 and oral formant bandwidths from F1 to F5 were calculated as means for the internal 50% duration of the consonant). The acoustic differences between [β̞] and [ʋ] were not particularly clear, perhaps with the exception of oral formant bandwidth values. The average acoustic characteristics of the stimuli are consistent with previous studies and the study presented here (see Chapter 4), but F2 values are considerably lower (for both [β̞] and [ʋ]).

Table 5.1. Averaged acoustic characteristics for the approximant consonants from nonsense words, used in the natural identification task.

	Bilabial		Labiodental	
	[ˈsa.β̞a]	[ˈga.β̞a]	[ˈsa.ʋa]	[ˈga.ʋa]
Duration (ms)	64.0	61.0	62.1	59.9
Intensity (dB)	67.3	68.6	66.4	65.3
f_0 (Hz)	136	137	131	132
F1 (Hz)	438	450	453	447
F2 (Hz)	1161	1172	1161	1191
F3 (Hz)	2524	2524	2531	2507
F4 (Hz)	3639	3590	3658	3608
F5 (Hz)	NA	NA	NA	4308
$F1_{bw}$ (Hz)	198	192	243	234
$F2_{bw}$ (Hz)	94	88	100	123
$F3_{bw}$ (Hz)	222	152	298	388
$F4_{bw}$ (Hz)	223	246	207	408
$F5_{bw}$ (Hz)	NA	NA	NA	NA

Out of these 32 stimuli, one from each category was selected as the best exemplar to serve as reference for the endpoint of synthetic continua, to be used in identification and discrimination tasks (task continua: from [ˈsa.β̞a] to [ˈsa.ʋa]; practice continua: from [ˈga.β̞a] to [ˈga.ʋa]). Care was taken that the bilabial to labiodental contrast was maximized in these pairs, as judged by auditory analyses conducted by the author. Waveforms and spectrograms for these stimuli can be seen in Figure 5.1. Stills for the point of maximum constriction for each approximant consonant are provided in Figure 5.2.

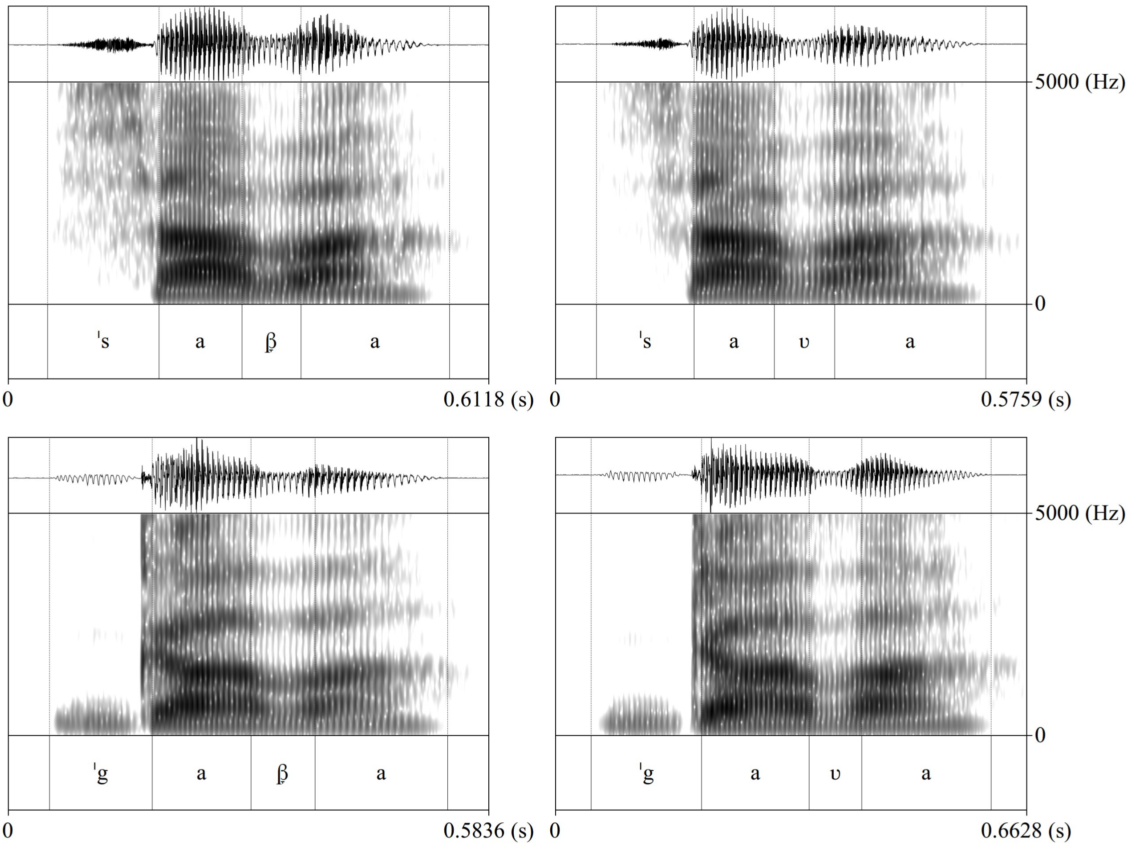

Figure 5.1. Spectrograms and waveforms for four natural nonsense words: top left panel “saba” [ˈsa.β̞a], top right “sava” [ˈsa.ʋa], bottom left “gaba” [ˈga.β̞a] and bottom right “gava” [ˈga.ʋa]. Nonsense words were segmented manually following visual and auditory inspection of the signals.

Figure 5.2. Still frames for the points of maximum articulatory constriction in the approximant consonants for four nonsense words: top left panel “saba” [ˈsa.β̞a], top right “sava” [ˈsa.ʋa], bottom left “gaba” [ˈga.β̞a], and bottom right “gava” [ˈga.ʋa]. A bilabial constriction can be seen in the left-hand side panels, whereas a labiodental constriction is observed in the right-hand side panels.

The selected recordings were manually excised in *Praat* (Boersma & Weenink, 2015). Approximant consonants and surrounding vowels for both endpoints were segmented with the aid of waveform, spectrogram and auditory inspection of the signals (see Figure 5.1). Acoustic models for the approximant consonants and surrounding vowels were built by extracting pitch, intensity, oral formants from F1 to F5 (maximum formant set to 5000 Hz for a male speaker), and oral formant bandwidth measurements from F1 to F5. The duration of each minimal pair was homogenized to that of the shorter consonant ([ʋ]). Each model consisted of 200 samples equally distributed along the time domain. KlattGrid objects were built and populated for the VCV bilabial to labiodental endpoints, as well as 7 intermediate equally-distanced steps. These objects were then synthesized to sounds using Klatt synthesis (Klatt & Klatt, 1990; Weenink, 2009) and spliced back with a 10 ms overlap to the onset consonant (/s/ or /g/). Mean intensity for all stimuli was scaled to 70 dB. The resulting synthetic endpoints for both continua can be seen in Figure 5.3.

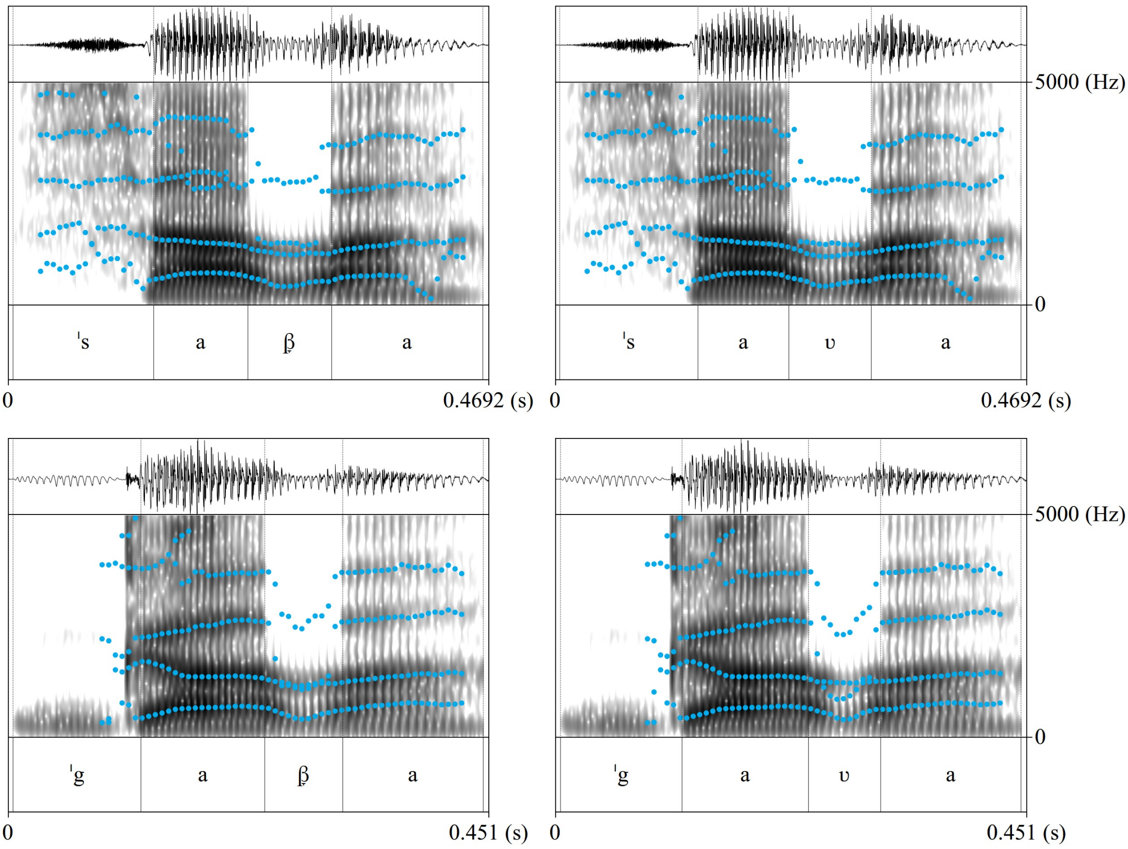

Figure 5.3. Spectrograms, waveforms and formant trajectories for the four synthetic target stimuli: top left panel “saba” [ˈsa.β̞a], top right “sava” [ˈsa.ʋa], bottom left “gaba” [ˈga.β̞a], and bottom right corner “gava” [ˈga.ʋa]. Formant trajectories for the first 5 formants are shown on top of the spectrograms, in blue.

The acoustic characteristics of the natural approximant consonants [β̞] and [ʋ] used to build the synthetic continuum from [ˈsa.β̞a] to [ˈsa.ʋa], as well as the acoustic characteristics from the synthesized consonants, can be found in Table 5.2. Besides duration and bandwidths for F1, F3 and F4, most acoustic variables from the natural stimuli display similar values. For the synthetic continuum, while the range of variation between [β̞] and [ʋ] was more or less preserved (duration and intensity were homogenized), it was clear that the synthesis process affected the relative position of some acoustic variables in their scales. This is particularly clear in oral formants from F3 to F5 and in all bandwidths. Given that a natural pair in which the contrast between [β̞] and [ʋ] was particularly clear, the stimuli from Table 5.2 are longer than the average approximant consonant of /b/. F1 values are slightly higher than those reported by the literature and those observed in the production study (see Chapter 4). F2 values are closer in these stimuli to those reported in the literature.

Table 5.2. Summary of the acoustic characteristics of the natural reference approximant consonants and of the synthetic continuum created from them (the first step based on [β̞] and the last one on [ʋ]).

	Natural		**Synthetic**								
	[β̞]	[ʋ]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Duration (ms)	96.0	91.3	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0	91.0
Intensity (dB)	59.0	57.3	66.5	66.3	66.1	65.9	65.7	65.5	65.3	65.2	65.0
f_0 (Hz)	140	137	139	139	139	137	137	137	137	137	136
F1 (Hz)	553	550	560	563	565	562	560	563	564	564	564
F2 (Hz)	1226	1217	1217	1218	1215	1212	1212	1208	1207	1206	1204
F3 (Hz)	2554	2490	2066	2068	2068	2069	2069	2072	2072	2070	2070
F4 (Hz)	3573	3561	3250	3252	3252	3254	3253	3254	3255	3255	3256
F5 (Hz)	4043	4378	4492	4494	4498	4503	4143	4143	4143	4143	4143
$F1_{bw}$ (Hz)	123	136	97	98	101	103	102	102	103	105	106
$F2_{bw}$ (Hz)	105	103	104	104	109	107	110	112	114	116	116
$F3_{bw}$ (Hz)	427	525	717	715	714	712	710	708	707	704	702
$F4_{bw}$ (Hz)	548	698	813	813	813	813	814	815	813	813	814
$F5_{bw}$ (Hz)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

5.3.3. Procedure

The experimental sessions were all conducted in the Phonetics Laboratory of Pontificia Universidad Católica, in Santiago, Chile, by a trained phonetician different from the author. Participants were seated in a quiet room in front of a computer which presented them with the experimental interface and stimuli. The participants' responses were registered using the computer's mouse and stored digitally. Sennheiser HD 201 headphones were used. The experimental interface was programmed and presented in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), a cross-platform experiment builder, managing Python experimental packages.

Before each experimental session, participants completed a short volume calibration procedure. A sequence of 6 Spanish words of varying length and interspaced with 300 ms silences was presented (*agendar*, “to schedule”; *vaporizar*, “to vaporize”; *fotografiar*, “to take pictures”; *cubo*, “cube”; *corrector*, “correction pen”; and *encuadernar*, “to bind”). These words were recorded by the author and their intensity was normalized to 70 dB. On hearing the sequence, participants had to indicate whether they were hearing the stimuli clearly and comfortably. If this was not the case, the volume was adjusted iteratively until listeners reported that it was set appropriately.

Identification of natural stimuli

At the start of the experimental session, participants were shown orthographic transcriptions of the nonsense words (“saba”, “sava”,“gaba” and “gava”), and were told to assume that they were common nouns. Participants were then presented with the stimuli and asked to identify the nonsense word between two options provided as buttons on the screen, which contained an orthographic transcription of the minimal pair relevant to the stimuli (“saba” and “sava”, or “gaba” and “gava”). The order of the response buttons was counterbalanced across participants. Each nonsense word had 8 different realizations. One token from each nonsense category was used twice in a short randomized practice session. For the task, the remaining 7 stimuli from each of the 4 nonsense words were presented 10 times, in a randomized order, amounting to 280 trials.

Identification of synthetic stimuli

The initial procedures and preparation were the same as in the natural speech condition. Participants completed a short practice session with a 9 step continuum from [ˈga.β̞a] to [ˈga.ʋa], presented in a randomized order. For the experimental task, participants were exposed to a 9 step continuum from [ˈsa.β̞a] to [ˈsa.ʋa], which was repeated 30 times in a randomized order, amounting to 270 trials.

Discrimination

Perception of the stimuli was tested in an ABX discrimination task with 6 stimulus pairs with a 2 step inter-stimulus distance (pairs: 1-4, 2-5, 3-6, 4-7, 5-8, 6-9). A 300 ms silence was inserted between “A” and “B” and a 500 ms silence between “B” and “X”. All possible permutations for the ABX design were included (“ABA”, “ABB”, “BAA” and “BAB”). Before the task took place, the ABX format was explained to participants. They were told that the first two elements of each trial (“A” and “B”) would always be different, despite being potentially very similar, and that the third element (“X”) would correspond to either the first or second one (these instructions apply to all permutations of the ABX format). Participants entered their responses using two buttons on the screen and a mouse. Participants completed two short practice sessions. The first practice session aimed to train the ABX task, using natural stimuli which maximized the contrast between [ˈsa.β̞a] and [ˈsa.ʋa], or between [ˈga.β̞a] and [ˈga.ʋa] (see Figure 5.4). Participants completed 8 trials, including both nonsense word pairs and several permutations of the internal elements of a trial.

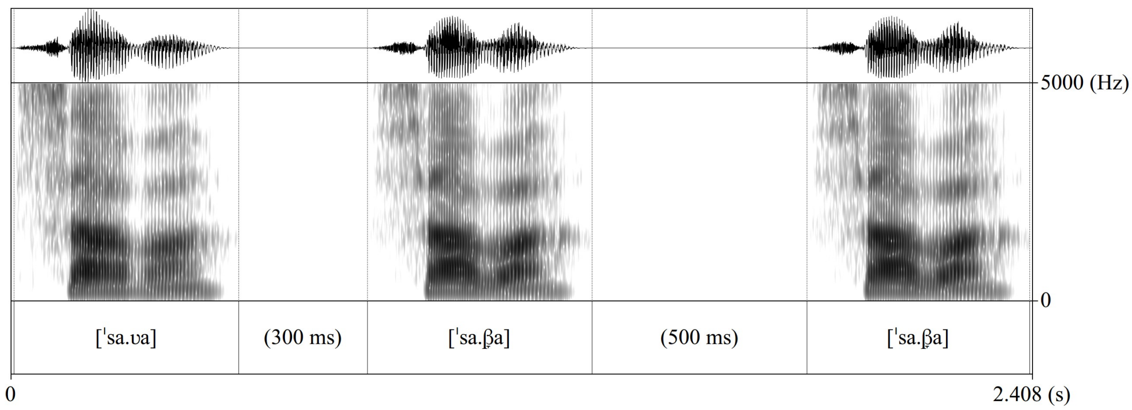

Figure 5.4. Waveform and spectrogram for a natural stimuli ABX trial (ABB), presented in the first practice session.

The second practice session consisted of listeners being presented with synthetic stimulus pairs taken from a continuum from [ˈga.β̞a] to [ˈga.ʋa]. Participants completed 8 trials in a randomized order, including different permutations of the order of the internal elements of each trial. This practice session was comparable in difficulty to the main task. For the main task, 6 pairings created from the task continuum from [ˈsa.β̞a] to [ˈsa.ʋa] were randomly presented 10 times in 4 internal permutations (ABA, ABB, BAA, BAB), amounting to 240 trials. See Figure 5.5 for an example.

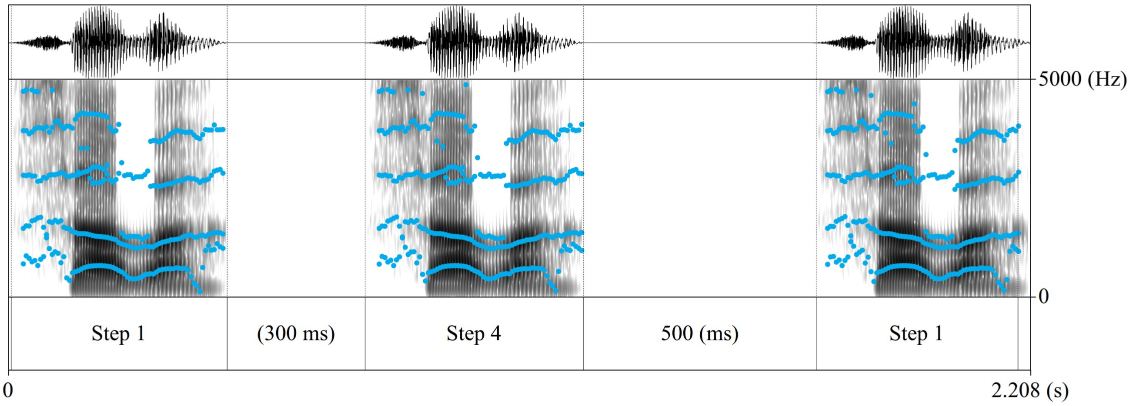

Figure 5.5. Waveform, spectrogram and formant trajectories for a synthetic ABX trial for the continuum from [ˈsa.β̞a] to [ˈsa.ʋa], presented in the experimental discrimination task. Formant contours are shown in blue.

5.4. Results

5.4.1. Identification of natural stimuli

The results for the identification of natural stimuli showed a trend for listeners to identify [β̞] as [β̞] more than as [ʋ] (see left-hand side panel from Figure 5.6), which suggests that, to some extent at least, listeners were able to identify [β̞] and [ʋ] as separate phonetic categories in natural stimuli. However, the results clustered around chance level and there was a large amount of overlap between categories. Moreover, inspection of individual variability (see right-hand side panel, Figure 5.6) showed that most listeners were not able to identify [β̞] and [ʋ] consistently, as suggested by a majority of lines near to the horizontal position, and some speakers identifying these categories in a direction opposite to the acoustic evidence.

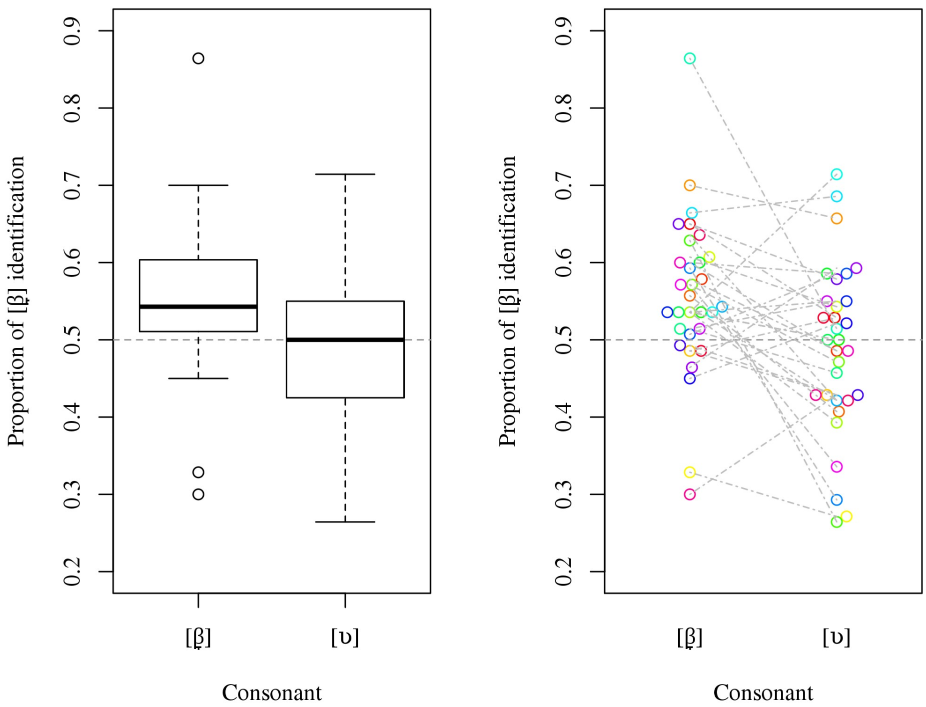

Figure 5.6. Boxplot and bee-swarm for natural stimuli [β̞] identification proportion for [β̞] and [ʋ] (n = 8400), averaged by subject (n = 30). On the bee-swarm, listeners are colour-coded and linked to themselves by dashed lines.

A generalized linear mixed model analysis (GLMM) was conducted on the results of natural stimuli identification to explore its statistical association to place of articulation (all subsequent GLMM analyses to be found in the following chapters will be conducted following the same general procedure). The analysis was carried out in *R* (R Core Team, 2013) using the *glmer* function from the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014), set for a binomial family and using a *logit* link function. For model selection, a null baseline model was fitted with natural identification results as the dependent variable, and participant and stimuli as random factors. Alternative more complex models were then created by adding a main effect and random slopes, and compared using the *anova* function until the best fitting model was found as judged by lower Akaike information criterion values (AIC; Akaike, 1998), lower Bayesian information criterion values (BIC; Schwarz, 1978), and the statistical significance of the differences observed between the compared models, provided by a chi-squared analysis on the residuals. The best fitting model for this analysis included natural identification results as dependent variable, place of articulation as main effect, and subject and stimuli as random factors. The assumption of normality for the residuals from this model was assessed via histograms and quantile-quantile plots; deviations from normality were observed, and thus the results of this analysis have to be interpreted cautiously. Wald chi-square tests (Type II) were calculated using the *Anova* function from the *car* package (Fox & Weisberg, 2011) to obtain confidence estimates for the main effects and interactions from the best fitting model. The results showed a significant main effect of place of articulation on the natural identification results ($\chi^2(1) = 4.8147$, $p < 0.05$).

5.4.2. Identification of synthetic stimuli

The results for synthetic stimuli identification for a continuum from [β̞] to [ʋ] showed no effect of place of articulation in identification (see Figure 5.7). Results for all levels of the continuum centred around chance level, with no visible deviations from this trend.

A GLMM analysis was conducted on the results of synthetic stimuli identification to explore the statistical significance of stimulus level on identification results. Adding stimulus level as a main effect during the model selection stage failed to improve its fit, as judged by lower AIC and BIC values, and the statistical significance of the differences observed between the compared models, provided by a chi-squared analysis on the residuals. In summary, no main effect of stimulus level was found.

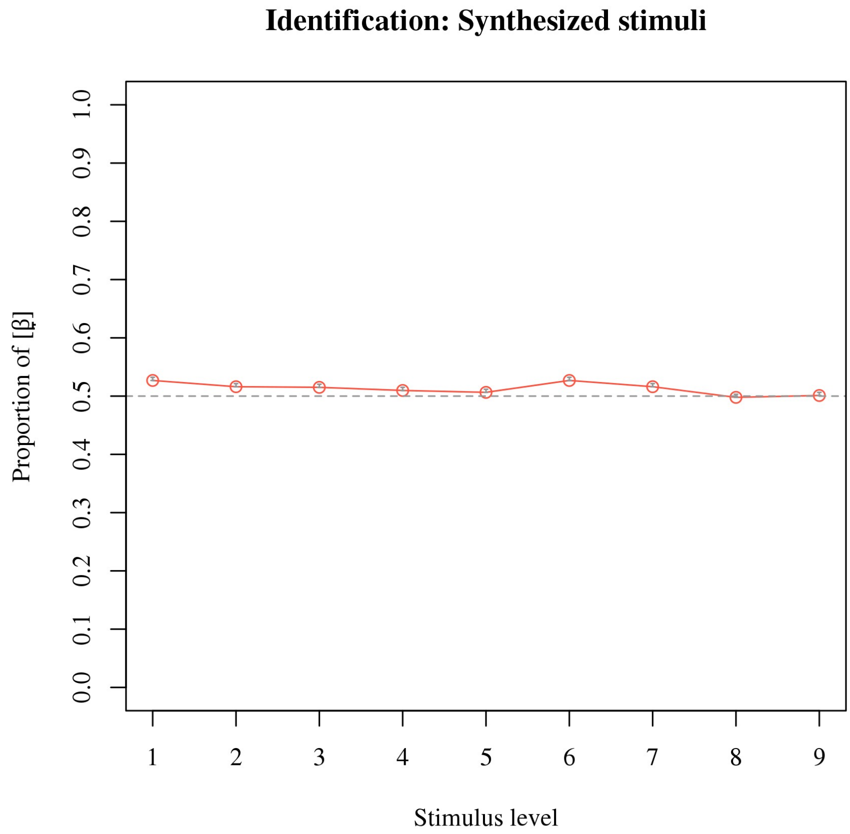

Figure 5.7. Average proportion of bilabial perception as a function of stimulus level for a 9 step continuum from [ˈsa.β̞a] to [ˈsa.ʋa] (n = 8100). 95% confidence interval bars are included.

5.4.3. Discrimination

The results for the discrimination of stimulus pairs taken from a [β̞] to [ʋ] continuum failed to show any discrimination sensitivity peak (see Figure 5.8). Like in the results for identification, responses centre around chance level, with no clear deviations from this general pattern.

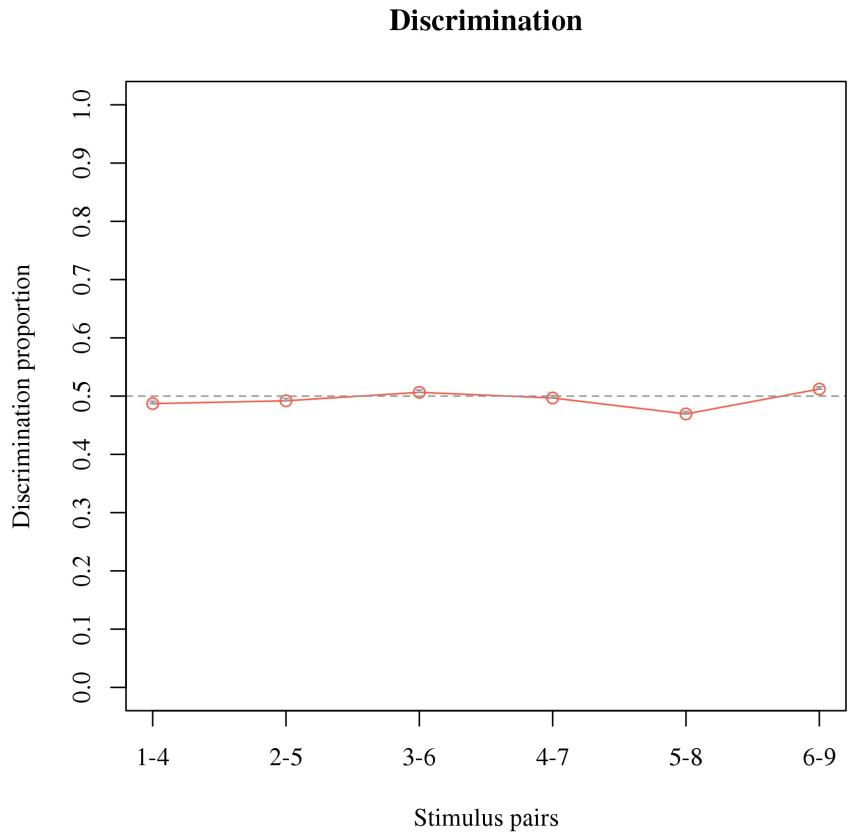

Figure 5.8. Average proportion of discrimination as a function of stimulus level pair (two-step inter-stimulus distance), taken from a 9 step continuum from [ˈsa.β̞a] to [ˈsa.ʋa] (*n* = 7200). 95% confidence interval bars are included.

A GLMM analysis was conducted on the results of synthetic stimuli identification to explore the statistical significance of stimulus pair on identification. Once more, adding stimulus pair as a main effect during the model selection stage failed to improve its fit, and thus it can be concluded that no main effect of stimulus level was present in the results.

5.5. Discussion

Until recently, research precedents dismissed the existence of labiodental variants of /b/ in Chilean Spanish. Despite some efforts to characterize their variation, the scope of their distributions and their relation to orthography (e.g., Sadowsky, 2010; Vergara Fernández, 2013), nothing is known about the acoustic differences between [β̞] and [ʋ], or about how listeners perceive these variants. This study set out to explore whether native speakers of Chilean Spanish are able to identify and discriminate bilabial and labiodental approximant consonants. This was done mainly to find out whether the acoustic differences between [β̞] and [ʋ] are perceived categorically, which would have direct implications in the design and interpretation of perception studies for /b/ (as those presented in Chapter 6 and Chapter 7). Three experiments were conducted: an identification task with natural stimuli, an identification task with synthesized stimuli and a discrimination task in an ABX format using synthetic stimulus pairs.

In the case of the first experiment, identification with natural stimuli, a main effect of place of articulation was found for the proportion of [β̞] responses: listeners showed a small tendency to identify each segment as coming from the corresponding place of articulation. However, most responses clustered around chance level, and both levels of place of articulation showed a large degree of overlap. Moreover, inspection of individual variation revealed that only a minority of listeners were able to identify [β̞] and [ʋ] reliably. These results seem to indicate that listeners might be sensitive to small acoustic differences in the continuum from [β̞] and [ʋ]. However, given that these differences were very small (see Table 5.1), it remains possible that listeners were paying attention to phonetic cues encoded, for example, in the immediate phonetic context (e.g., surrounding vowels), or that prosodic information was driving their responses. If listeners were systematically sensitive to differences between [β̞] and [ʋ], and both segments were perceived as different categories, minimal overlap between the two levels would have been expected.

The results from the identification task with synthetic stimuli were at chance level along the entire continuum. No main effect of stimulus level was detected in the statistical analyses. In summary, there was no evidence that listeners were able to identify synthetic stimuli from [β̞] and [ʋ]. In the case of the discrimination task, no sensitivity peaks were observed in any stimulus pair. Instead, all results were at chance level. No main effect of stimulus pair was found in the statistical analyses. Taken together, identification and discrimination tasks failed to provide any evidence of categorical discrimination for synthetic stimuli modelling a continuum from [β̞] to [ʋ].

The contradicting results of the natural identification task and identification and discrimination tasks with synthetic stimuli require some attention. Firstly, although a main effect of place of articulation was found in the natural identification task, results were far from conclusive, given the complete overlap between categories, the clustering of the results around chance level and the conflicting evidence from individual variability. Secondly, it is unlikely that the differing results from the natural identification task and the tasks using synthetic stimuli are due to the synthetic stimuli failing to represent the acoustic characteristics from natural [β̞] and [ʋ], or that the differences between [β̞] and [ʋ] were larger in natural stimuli. Inspection of Table 5.3 reveals that the average differences between [β̞] and [ʋ] in the acoustic variables from all stimuli (natural, natural reference for synthetic and synthetic) are within the same order of magnitude. Also, at least for those variables known to be particularly relevant to characterize vocoids in Spanish –duration, F1 and F2–, the differences are below known perception thresholds (e.g., Nooteboom & Doodeman, 1980).

Table 5.3. Averaged acoustic differences between [β̞] and [ʋ] in: (a) the natural stimuli used in the natural identification task, (b) the natural stimuli used as reference to build the synthetic stimuli, and (c) the endpoints of the [β̞] to [ʋ] synthetic continuum.

	Average acoustic differences		
	Natural stimuli	**Reference synthetic**	**Synthetic stimuli**
Duration (ms)	1.5	4.7	0.0
Intensity (dB)	2.1	1.7	1.5
f_0 (Hz)	5.0	3.0	3.0
F1 (Hz)	6.0	3.0	4.0
F2 (Hz)	9.5	9.0	13.0
F3 (Hz)	5.0	64.0	4.0
F4 (Hz)	18.5	12.0	6.0
F5 (Hz)	NA	335.0	349.0
$F1_{bw}$ (Hz)	43.5	13.0	9.0
$F2_{bw}$ (Hz)	20.5	2.0	12.0
$F3_{bw}$ (Hz)	156.0	98.0	15.0
$F4_{bw}$ (Hz)	73.0	150.0	1.0
$F5_{bw}$ (Hz)	NA	NA	NA

Although it could be argued that the relevant acoustic differences between [β̞] and [ʋ] were located in higher oral formants and their bandwidths, and that the synthetic stimuli failed to convey these differences[15], a more likely explanation is that listeners had considerably more acoustic cues available in the case of the natural stimuli, and perhaps some of them (e.g., duration of the neighbouring vowels) drove the main effect observed in the data. This is to be expected since a synthetic stimulus only models acoustic reality, which is degraded in the synthesis process at the same time that naturalness is compromised.

Taken together, these results indicate that listeners are not able to identify or discriminate [β̞] and [ʋ]. No evidence of categorical perception was observed for these segments, and in consequence it is unlikely that Chilean Spanish speakers encode any linguistic or extra-linguistic information using this contrast, unless that information is conveyed via visual cues not tested here. Given these results, no attempt will be made to control for place of articulation of variants of /b/ in the design of perception experiments in Chapter 6 and Chapter 7.

15 For example, it is well-known that F3 is a correlate of lip-rounding in some languages (e.g., Curtin, Fennell & Escudero, 2009). In the articulation of both [β̞] to [ʋ] there is intervention from the lips.

Chapter 6
Recovery, lexical effects and lexical access in /b d g/

6.1. Introduction

Under normal circumstances, that is, during spontaneous speech, listeners are often required to achieve lexical access for word forms for which they lack sufficient acoustic evidence (Mitterer & Ernestus, 2006). Highly lenited and elided variants are indeed the norm in conversational speech (Ingram, 1989; Fosler-Lussier & Morgan, 1999; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Johnson, 2004; Janse, Nooteboom, & Quené, 2007; Torreira & Ernestus, 2011; Brown, 2011), and one consequence of this is that the acoustic variables cueing for some segments are often poorly represented in the signal or completely absent from it. Despite these challenges, communication does not seem to be hindered by elision or lenition (Ernestus, 2014); on the contrary, listeners are capable of interpreting the perceptual input most of the time.

The fact that listeners are able to deal with unreliable acoustic evidence raises a series of questions pertaining to the sources of information and strategies that listeners employ to attain lexical access. One important strategy to aid perception is phonological recovery, whereby underlying representations for missing segments are formed despite lacking full prelexical support, provided certain conditions are met (Samuel, 1981a; Samuel, 1987; Samuel, 1996). For example, studies have shown that listeners can resort to coarticulatory cues to aid perception, taking advantage of coarticulatory information from segments preceding or following missing or masked ones in order to recover them (Yeni-Komshian & Soli, 1981; Repp, 1983). Listeners can also employ semantic and syntactic cues to achieve phonological recovery, especially when the acoustic cues are unreliable. For instance, studies on highly lenited forms have provided evidence showing that listeners are able to recover missing segments when additional semantic and syntactic contexts are provided (Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder. & Baayen, 2004; Mitterer & Ernestus, 2006).

Describing the conditions required for listeners to recover missing units from lenited and elided word forms is relevant because it has direct consequences for models of lexical access and speech perception. For instance, proponents of episodic models of speech perception such as LAFS (Klatt, 1979; 1989) and Minerva 2 (Hintzman, 1984, 1986; Goldinger, 1998) claim that episodic models are able to account for lexical access of lenited forms given that reduced word forms have their own episodic representations in long-term memory, which are activated to match the acoustic input when required, without having to resort to intermediate abstract representations or phonological recovery. However, episodic models have been challenged by evidence showing that listeners are unable to recover highly lenited word forms unless additional context is provided (e.g., Ernestus et al., 2002; Kemps et al., 2004). Pure bottom-up abstractionist models such as Cohort (e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987, 1989) and Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997), with no top-down lexical feedback to prelexical stages of speech processing, are also challenged by phonological recovery evidence, because these models require intermediate abstract representation to be formulated based exclusively on reliable acoustic evidence, which is not available in highly lenited forms. Interactive models of lexical access such as TRACE (McClelland & Elman, 1986a, 1986b), and hybrid models such as Goldinger's CLS (Goldinger, 2007), Pierrehumbert's ED (Pierrehumbert, 2001, 2002) and POLYSP (Hawkins & Smith, 2001; Hawkins, 2003), seem to fare better at accommodating experimental evidence in favour of recovery by positing that top-down feedback from the lexical level can inform prelexical stages of speech processing.

Although the hypothesis of phonological recovery itself is still debatable, particularly given that underlying representations are not an assumption of all models of lexical access, the fact remains that listeners are overwhelmingly successful at dealing with highly lenited forms when complementary sources of information compensate for the lack of reliable acoustic evidence. The way in which these additional sources of information contribute to achieve lexical access has been the focus of research on highly lenited forms from conversational speech. For example, Ernestus et al. (2002) presented word forms that varied in their degree of lenition and carefully controlled the amount of acoustic and semantic cues available to the listener. The results demonstrated that listeners were able to recognize highly reduced word forms when phonetic, semantic and syntactic contexts facilitated lexical access. Another experiment by Kemps et al. (2004), showed that listeners were able to recover underlying /l/ from highly reduced instances of the Dutch suffix "-(e)lijk" [(ə)lək] in phoneme monitoring tasks, but only when the reduced suffixes were presented in a context of several words. Mitterer and Ernestus (2006) also conducted a series of perception experiments in which they presented synthetic instances of Dutch /t/ in coda position with varying degrees of lenition in several phonetic contexts, and in several semantic and syntactic contexts; their results showed that listeners utilized both bottom-up (phonological context and sub-phonemic detail) and top-down sources of information (lexical status) in order to attain lexical access.

Chilean Spanish spirant approximant variants of /b d g/ provide a novel testing ground to explore some of these issues surrounding the nature of phonological recovery, and how listeners weight different cues depending on their availability. This is because [β̞], [ð̞] and [ɣ̞] display natural continua of realizations from open approximants to elided variants (for details, see Chapter 3 and Chapter 4). While most of the research on highly lenited forms presented categorical increments of acoustic information to listeners (Ernestus et al., 2002; Kemps et al., 2004), spirant approximants of Spanish allow the amount of acoustic evidence available in perception to be carefully controlled. Experiments can thus be designed in which increasing amounts of acoustic detail can be presented, enabling exploration of the role of fine-grained phonetic detail in the perception and phonological recovery of lenited forms.

Another important advantage of this particular form of variation is that, for some minimal pairs, both the presence of an approximant consonant and its absence constitute legal words. For example, eliding the approximant consonant from the word *boga* [ˈbo.ɣ̞a] ("fashionable" or "trendy") renders [ˈbo.a], which can be interpreted by the listener as *boga*, with lenited /b/, or as *boa* ("boa constrictor"). Minimal pairs such as these allow the creation of ecologically valid continua from consonant presence to absence, while at the same time controlling for some lexical effects on speech perception, which would otherwise bias perception towards words as opposed to nonsense words (Ganong, 1980). Crucially, how listeners process these minimal pairs offers a transparent way to determine whether they actually perform phonological recovery, since the underlying phonological unit is the only difference between the two items.

Unlike previous research, this study will use synthetic continua from consonant presence to absence in several informational conditions: minimal phonetic context, word-level context and semantically primed word-level context (in two directions). For all continua, both ends are Spanish words, and each interpretation of a continuum will also be semantically primed in separate conditions. The first condition, minimal phonetic context, establishes an auditory baseline of perception for each consonant, against which the other three conditions will be compared. The second condition, word-level context, will present the same continua embedded in a lexical context. The last informational level consists of two separate conditions: semantic priming of the lexical item containing the full approximant consonant (e.g., *boga*), and of the item with no approximant (e.g., *boa*). These last two conditions aim to bias the perception of continua towards the primed item, potentially enabling further feedback from the lexical level to prelexical processing when compared to the word-level condition. Continua will be presented in a modified version of the phoneme monitoring task, and in traditional identification and discrimination tasks. The phoneme monitoring tasks aims to obtain responses closer to auditory processing, in which lexical access is not mandatory[16]. The identification task will provide the listener with predetermined categories from which to choose an answer, and thus lexical level processing will be mandatory. Finally, a discrimination task will be paired with an identification task in order to explore categorical perception hypotheses (McQueen, 1996).

6.2. Aims

(a) Establish an auditory perceptual baseline for the perception of approximant consonants of /b d g/ in continua from approximant to elided variants.
(b) Determine the effect of increasing the number and type of acoustic (full phonetic context) and non-acoustic cues (full phonological context, minimal semantic context and semantic priming) in the perception of continua from approximant

16 I refer to this task as a modified version of the phoneme monitoring task because no reaction times will be collected (Newman & Dell, 1978; Frauenfelder & Segui, 1989; Titone, 1996).

to elided variants in /b d g/.

(c) Determine whether there is evidence of phonological recovery in the perception of /b d g/ in any of the experimental conditions.

(d) Determine the similarities and differences in the perception of continua from approximant to elided variants for the three phonological categories /b/, /d/ and /g/.

(e) Interpret the results in light of lexical access models and models of speech perception.

6.3. Methods

6.3.1. Participants

Sixty one native monolingual Chilean Spanish speakers (mean age 21.1 years; 42 females and 19 males) took part in the experiments, which consisted of two sessions lasting around 1 hour each and taking place on different days. Participants were undergraduate students, residents of Santiago (*n* = 48), Concepción (*n* = 11) and two other large Chilean urban centres. Participants received an information sheet prior to the experiment, and were required to give informed consent and fill in a short questionnaire before the start of the first experimental session. None of the participants reported having any cognitive, hearing, language or speech impairment. Participants were paid for their participation.

6.3.2. Stimuli

Several minimal pairs such as *boga* [ˈbo.ɣ̞a] (“fashionable” or “trendy”) and *boa* [ˈbo.a] (“boa constrictor”), in which eliding an intervocalic approximant consonant from the first word results in a different lexical unit, were identified for /b/, /d/ and /g/. These pairs and each item's relative lexical frequency expressed as instances per million words were extracted from the non-lemmatized lexical Spanish frequency list CREA (Real Academia Española, 2014). Minimal pairs in which both items shared the same lemma (e.g., *atraída* [a.tɾa.ˈi.ð̞a], "he/she/it was attracted", versus *atraía* [a.tɾa.ˈi.a], "he/she/it attracted") were excluded, as well as particularly unusual word forms. Lexical frequency for the members of a minimal pair was homogenized so that the relative lexical frequency for the most frequent item did not exceed more than two times that of the less frequent item. Although complete lexical frequency homogeneity would have been desirable, any stricter criteria failed to render enough usable minimal pairs.

Five semantic associates were selected for each lexical item to serve as primes. Care was taken to ensure that no prime had the same morphological structure as its corresponding target word. The primes were submitted to an online word association task in which 20 monolingual native Chilean Spanish speakers quantified the strength of the association between the target (e.g., *dudo*, "to doubt") and a given prime (e.g., *titubear*, "to hesitate") by means of a Likert scale ranging from no association at 0 to maximum association at 7. The mean age of the participants was 26 years (14 females and 6 males). All participants gave informed consent and were compensated for their participation. The primes with the highest semantic association index were selected for each target word. Care was taken to ensure that the two associates for each minimal pair had a similarly strong associate. The mean association strength for the selected primes was 5.87, and the average association difference within minimal pairs was 0.39.

Several instances of all target words and their selected associates were recorded by the author, a monolingual native Chilean Spanish speaker, in a sound-isolated booth. A Rode NT1A condenser microphone was used, along with an RME Fireface UC interface connected to a PC. Recordings were made in TotalMix mixer at a frequency of 44100 Hz and 16 bit depth. The recordings were filtered with a Hann band-stop filter from 0 to 60 Hz. All words were then excised manually using Praat (Boersma & Weenink, 2015). For both the full form (e.g., *dudo*) and the elided targets (e.g., *dúo*), the segments of interest were segmented manually into TextGrids using visual cues from waveforms and spectrograms, and through auditory inspection of the signal (see Figure 6.1).

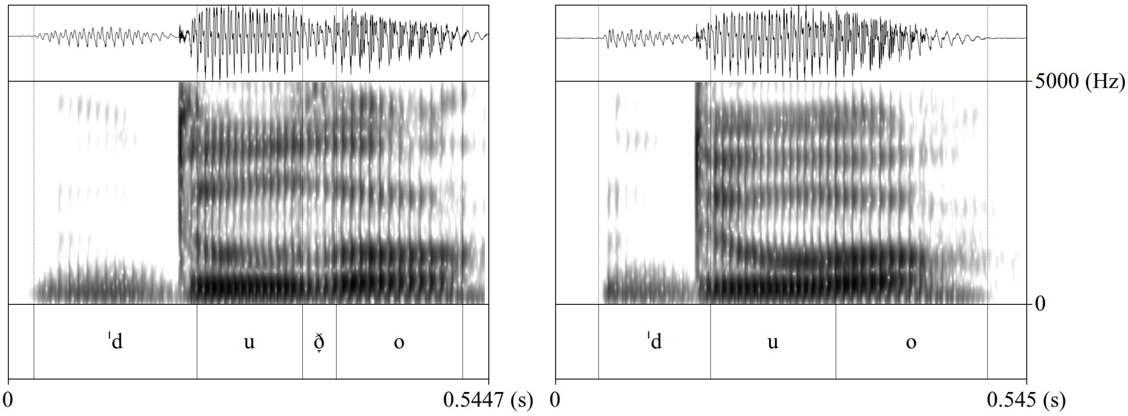

Figure 6.1. Left-hand side panel: Waveform and spectrogram for an instance of *dudo* [ˈdu.ð̞o], a full form target. The intervocalic approximant consonant [ð̞] is visible in both the waveform and spectrogram. Right-hand side panel: Waveform and spectrogram for an instance of *dúo* [ˈdu.o], the elided target. The two vowels transition into each other and the approximant consonant is fully elided.

An acoustic model was built for the approximant consonants and their neighbouring segments (e.g., for [ˈu.ð̞o], from *dudo*), as well as for the corresponding elided counterparts (e.g., [ˈu.o], from *dúo*). To build these models, the time domain was divided into 100 equally distanced samples where f_0, intensity, oral formants from F1 to F3, and bandwidths for F1 to F3 were queried from acoustic objects in Praat. KlattGrid objects were created for each section and then populated with the acoustic parameters to match the acoustic models. Eight equally-distanced intermediate steps were created between the full approximant and elided targets. The resulting 10 steps were synthesized into sounds with a sampling frequency of 44100 Hz, using Klatt synthesis (Klatt & Klatt, 1990; Weenink, 2009). Examples of the endpoints for a continuum can be seen in Figure 6.2.

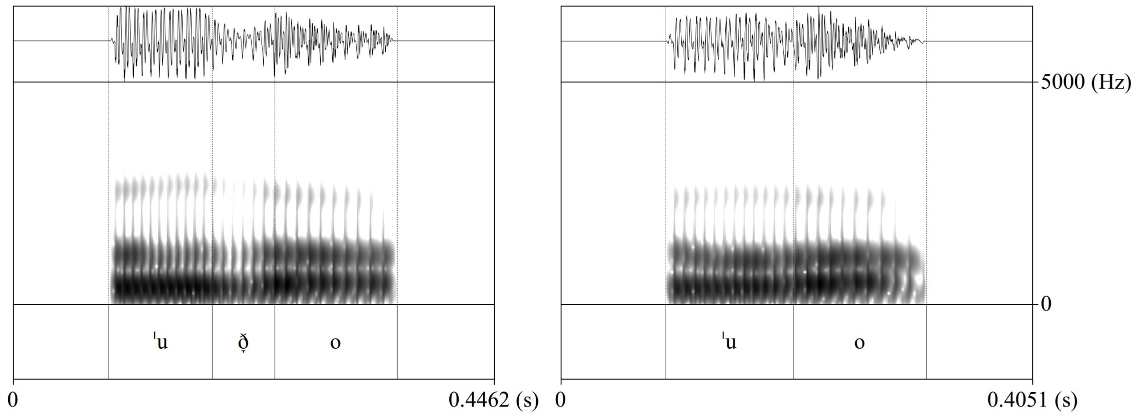

Figure 6.2. Waveform and spectrogram for the synthesized endpoints of the [ˈu.ð̞o] to [ˈu.o] continuum. The approximant consonant is visible between the vowels in the left-hand side panel, and is fully lenited in the right-hand side panel.

Two conditions – word and primed word – required the resulting synthetic sections to be spliced into a broader phonetic context. For these conditions, each synthetic VCV or VV section was spliced back with overlap to the remaining unaltered section taken from the original full approximant target. For example, in the case of the minimal pair *dudo* versus *dúo*, the word-initial [d] from the full form *dudo* was spliced back at the beginning of the 10 synthesized steps (see Figure 6.3 for an example). In those cases in which the vowel following the approximant consonant was located at the end of a word, a short fade-out was applied to avoid noticeable clicks from appearing.

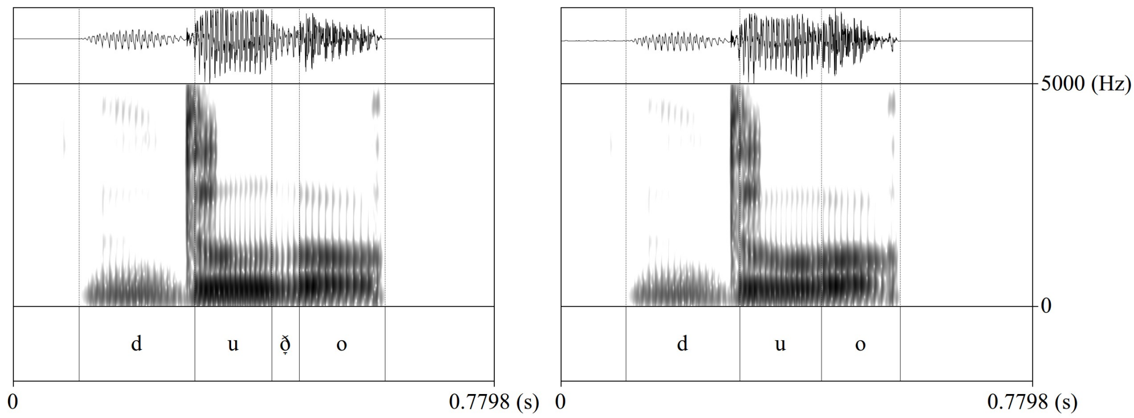

Figure 6.3. Waveforms and spectrograms for the synthesized endpoints for the [ˈdu.ð̞o] to [ˈdu.o] continuum. The approximant consonant is visible in the waveform and spectrogram in the left-hand side panel. In the right-hand side panel the consonant is fully elided.

In order to obtain stimuli that sounded as naturalistic as possible, the segmentation of the signals, the amount of overlap for splicing and the start of the optional fade-out were adjusted iteratively until a satisfactory version of each continuum was created, as judged by auditory inspections conducted by the author and another trained phonetician (McQueen, 1996). Afterwards, all stimuli including semantic associates were subjected to a Hann band-pass filter from 0 to 5000 Hz in order to match the maximum frequency and overall quality of the synthetic sections with the rest of the natural stimuli. Mean intensity was homogenized to 70 dB.

The best two continua were selected for approximants of /b/, /d/ and /g/, one for the tasks and another for the practice sessions. The continuum for the practice sessions for /b/ was created from the words *releva* ("to take over") [re.ˈle.β̞a] and *relea* ("to re-read") [re.ˈle.a]. The semantic prime for the full form *releva* was *reemplazar* ("to replace"), and for the elided endpoint *repasar* ("to recapitulate"). The continuum for the tasks for /b/ was created from the words *cubetazo* ("to hit with a bucket or its content") [ku.β̞e.ˈta.so] and *cuetazo* ("fire-cracker explosion") [kwe.ˈta.so]. The prime for the full form *cubetazo* was *balde* ("bucket") and the semantic prime for the elided endpoint was *explosión* ("explosion"). A summary of the main acoustic characteristics for the natural [β̞] from the full word *cubetazo,* and for the same consonant in the synthetic continuum can be found in Table 6.1. The [β̞] from the first synthetic step, [ˈu.β̞e], was, for the most part, equivalent to the natural consonant on which it was based[17]. As for the synthetic continuum, duration decreased gradually while intensity increased as the steps became closer to elided variants. Formant values also increased or decreased gradually to match the acoustic characteristics of the point where the surrounding vowels met.

The continuum for the practice sessions for /d/ was created from the words *callado* ("silent") [ka.ˈʝ̞a.ð̞o] and *Callao* ("Callao", the Peruvian port) [ka.ˈʝ̞a.o]. The semantic prime for the full form *callado* was *enmudecer* ("to silence"), and for the elided endpoint *puerto* ("port"). The continuum for the tasks for /d/ was created from the words *dudo* ("to doubt") [ˈdu.ð̞o] to the word *dúo* ("duet") [ˈdu.o]. The prime for the full form *dudo* was *titubear* ("to hesitate") and the semantic prime for the elided endpoint was *pareja* ("couple"). A summary of the main acoustic characteristics for the

17 The sequence [ˈu.β̞e] was extracted from the word [ku.β̞e.ˈta.so], whose lexical stress is located on the penultimate syllable. The stress mark included in [ˈu.β̞e] reflects a secondary stress.

Table 6.1. Summary of acoustic characteristics for the natural [β̞] approximant consonant from the full approximant target stimuli *cubetazo* ([ku.β̞e.ˈta.so]), and of the same consonant from the synthetic continuum between [ˈu.β̞e] and [we], created from the words *cubetazo* (step 1) and *cuetazo* (step 10), for the tasks involving /b/. Duration refers to the duration of the approximant consonant. Intensity to the minimum intensity after normalization for the VCV to VV section. Fundamental frequency, oral formant values from F1 to F3 and oral formant bandwidths from F1 to F3 are provided as means for the internal 50% duration of the consonant.

	[β̞]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Duration (ms)	47.0	44.8	37.3	30.5	24.3	18.8	13.8	9.4	5.7	2.5	0
Intensity (dB)	69.8	65.2	65.8	67.0	67.8	68.5	69.2	70.0	70.1	70.2	NA
f_0 (Hz)	85.4	86.4	86.9	87.4	88.0	88.5	89.1	89.4	89.7	90.0	NA
F1 (Hz)	327	332	337	344	364	365	367	373	388	390	NA
F2 (Hz)	963	948	1013	1068	1235	1299	1304	1324	1511	1534	NA
F3 (Hz)	2166	1996	1988	1985	2075	2112	2058	2043	2090	2070	NA
$F1_{bw}$ (Hz)	93	78	54	38	112	94	61	30	80	80	NA
$F2_{bw}$ (Hz)	159	159	132	99	323	315	267	137	415	489	NA
$F3_{bw}$ (Hz)	423	660	426	282	1464	1278	498	162	796	470	NA

natural intervocalic [ð̞] from *dudo* and for the synthetic continuum for /d/ can be found in Table 6.2. Duration decreased as a function of stimulus step, while minimum intensity increased as the elided unit was lenited and became more similar to the surrounding segments. Some differences were observed between the first synthetic step for the continuum and the natural [ð̞], particularly in F3 values and bandwidth from F1 to F3. These dissimilarities could correspond to actual differences in the formant constitution of these units or to artefacts originating from the FFT analyses. One possible explanation for the latter case would be that the natural approximant consonant had a richer harmonic and formant structure, and the FFT analysis was optimized for 5 formants in a 0 to 5000 Hz frequency space, while the synthetic samples were built with 3 oral formants and the analysis was optimized for 3 formants in the same frequency range. In any case, less reliable results are expected from measurements taken from very short stimuli.

Table 6.2. Summary of acoustic characteristics for the natural [ð̞] approximant consonant from the full approximant target stimuli *dudo* ([ˈdu.ð̞o]), and of the same consonant from the synthetic continuum between [ˈu.ð̞o] and [ˈu.o] created from the words *dudo* (step 1) and *dúo* (step 10), for the tasks involving /d/.

	[ð̞]	1	2	3	4	5	6	7	8	9	10
Duration (ms)	38.3	35.9	31.3	26.9	22.7	18.6	14.6	10.7	7.0	3.4	0
Intensity (dB)	70.5	64.6	65.1	65.8	66.4	66.9	67.4	67.9	69.9	70.1	NA
f_0 (Hz)	105.9	107.6	107.4	107.0	106.6	106.2	105.9	105.6	105.3	105.2	NA
F1 (Hz)	331	348	343	347	359	373	378	390	401	410	NA
F2 (Hz)	1070	1118	1053	1031	1040	1045	1020	1014	1010	1004	NA
F3 (Hz)	2650	3045	2603	2437	2507	2593	2384	2342	2359	2422	NA
$F1_{bw}$ (Hz)	94	142	84	38	83	124	44	35	41	60	NA
$F2_{bw}$ (Hz)	211	462	204	74	143	181	57	46	55	94	NA
$F3_{bw}$ (Hz)	165	2201	660	212	523	860	229	184	247	433	NA

The continuum for the practice sessions for /g/ was created from the words *mega* (“mega”) [ˈme.ɣ̞a] and *mea* (“to urinate”, informal) [ˈme.a]. The semantic prime for the full form *mega* was *grande* (“big” or “large”), and for the elided endpoint *orinar* (“to urinate”, formal). The continuum for the tasks for /g/ was created from the words *boga* (“fashionable”, “trendy”) [ˈbo.ɣ̞a] to the word *boa* (“boa constrictor”) [ˈbo.a]. The prime for the full form *boga* was *actualidad* (“presently” or “current”) and the semantic prime for the elided endpoint was *constrictor* (“constrictor”). A summary of the main acoustic characteristics for the natural [ɣ̞] from *boga* ([ˈbo.ɣ̞a]) and for the same consonant in the synthetic continuum for /g/ can be found in Table 6.3. Duration decreased and intensity increased as a function of continuum step, as the consonant becomes more lenited and more similar to the surrounding vowels. The formant values for the first step from the synthetic continuum are representative of those from the natural recording.

Absolute, relative frequencies and semantic association strength for all selected stimuli are provided in Table 6.4. Absolute and relative frequencies have been taken from the non-lemmatized lexical Spanish frequency list CREA (Real Academia Española, 2014). Relative frequency was calculated as items per million words. The restrictions imposed on the potential minimal pairs regarding phonetic context, relative lexical frequency differences, lemma differences and morphological structure differences resulted in the availability of only a few candidates, most of them with low

Table 6.3. Summary of acoustic characteristics for the natural [ɣ̞] approximant consonant from the full approximant target stimuli *boga* ([ˈbo.ɣ̞a]), and of the same consonant from the synthetic continuum between [ˈo.ɣ̞a] and [ˈo.a] created from the words *boga* (step 1) and *boa* (step 10), for the tasks involving /g/.

	[ɣ̞]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Duration (ms)	50	46.3	40.1	34.1	28.3	22.9	17.8	12.9	8.3	4	0
Intensity (dB)	66.3	57.1	58.7	59.9	60.8	61.8	63.3	65.0	67.1	69.0	NA
f_0 (Hz)	91.3	91.9	92.5	93.2	93.9	94.5	95.3	96.1	97.1	97.9	NA
F1 (Hz)	361	365	374	410	397	417	441	457	474	485	NA
F2 (Hz)	1323	1306	1297	1331	1214	1175	1142	1104	1081	1053	NA
F3 (Hz)	2210	2075	2162	3867	2164	2213	2277	2256	2290	2309	NA
$F1_{bw}$ (Hz)	59	22	43	229	37	53	73	33	24	24	NA
$F2_{bw}$ (Hz)	56	57	74	130	23	42	81	45	32	27	NA
$F3_{bw}$ (Hz)	197	197	431	NA	206	239	344	168	140	163	NA

absolute and relative lexical frequencies (cf. Stemberger & MacWhinney, 1986; Bybee, 2000).

Table 6.4. Absolute frequency (AF), relative frequency (RF) and semantic associate strength (AS) for all selected stimuli and primes of /b/, /d/ and /g/, for both tasks and practice.

Phoneme	**Task**	**Status**	**Word**	**Section**	**AF**	**RF**	**Prime**	**AS**
/b/	Practice	Full form	*releva*	[ˈe.β̞a]	54	0.35	*reemplazar*	5.65
		Elided form	*relea*	[ˈe.a]	30	0.19	*repasar*	5.75
	Task	Full form	*cubetazo*	[ˈu.β̞e]	1	0.00	*balde*	6.04
		Elided form	*cuetazo*	[we]	1	0.00	*explosión*	5.87
/d/	Practice	Full form	*callado*	[ˈa.ð̞o]	1258	8.24	*enmudecer*	6.00
		Elided form	*Callao*	[ˈa.o]	593	3.88	*puerto*	5.64
	Task	Full form	*dudo*	[ˈu.ð̞o]	1006	6.59	*titubear*	5.85
		Elided form	*dúo*	[ˈu.o]	1026	6.72	*pareja*	6.62
/g/	Practice	Full form	*mega*	[ˈe.ɣ̞a]	179	1.17	*grande*	6.24
		Elided form	*mea*	[ˈe.a]	278	1.82	*orinar*	6.83
	Task	Full form	*boga*	[ˈo.ɣ̞a]	432	2.83	*actualidad*	5.18
		Elided form	*boa*	[ˈo.a]	184	1.20	*constrictor*	4.87

Finding minimal pairs in which both members displayed similar lexical frequencies proved to be very difficult, at least partly due to the nature of lexical frequency itself, in which a low number of items have high frequencies, and a much larger number of items displays frequencies close to zero, in what resembles an exponential-logarithmic distribution (see Figure 6.4). Given that lexical frequency has been shown to have an effect on lexical access and language processing (Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977; Segui, Mehler, & Frauenfelder, 1982; Lindblom, 1990; Ellis, 2002), on the rate of diffusion of phonetic changes (Bybee, 2000, 2002), and on lexical effects on speech perception (Fox, 1984), homogenizing relative

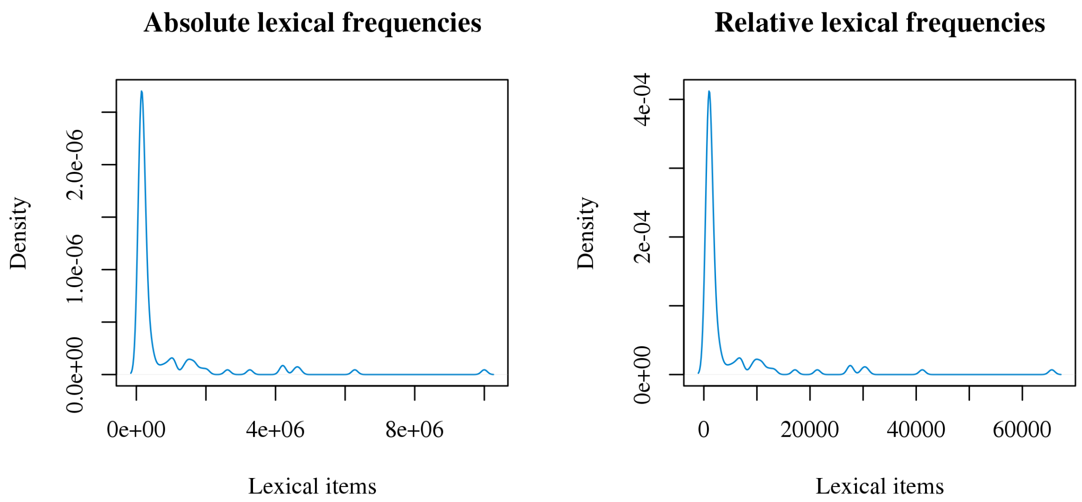

Figure 6.4. Kernel density plots for the absolute and relative lexical frequencies for the 100 more frequent tokens from the non-lemmatized lexical Spanish frequency list CREA (Real Academia Española, 2014).

lexical frequency to some extent was paramount for the perception experiments. As a consequence, some minimal pairs display very low lexical frequencies. There are also two potential problems with some minimal pairs. Firstly, in the case of /b/, the sequences [ˈe.β̞a] and [ˈu.β̞e] can actually be interpreted as the words *Eva* (/ˈe.ba/), “Eva”, and *hube* (/ˈu.be/), first person singular imperfect, indicative mood, from *haber*, “to be” or “to exist” (used only rarely, in compound verbs in very formal writing). However, the sequence [ˈe.β̞a] was only used in practice sessions, and in the case of [ˈu.β̞e] no evidence of a lexical effect was observed in the results of the segmental conditions in perception experiments (see “6.4. Results”). Secondly, the full form for /d/, *dudo*, contains a word-initial [d] that could prime [ð̞] when presented together. In order to address this, particular care was taken to remove all traces of the word-initial [d] from the segmental conditions, in which continua from [ˈu.ð̞o] to [ˈu.o] will be presented to participants.

6.3.3. General procedures

Data collection was conducted by 4 Linguistics graduate students from *Pontificia Universidad Católica de Chile* and *Universidad de Concepción*, formally trained in experimental phonetics. The main project and the purpose of the individual tasks were carefully explained to the interviewers. They were trained in all the procedures involved in the data collection, including the use of relevant software, hardware, ethics considerations and administrative tasks. Short pilot training sessions were completed by all interviewers prior to testing with participants. The experimental sessions took place in Santiago and Concepción. Participants were seated comfortably in a quiet room in front of a computer which presented them with the stimuli and experimental interface. The responses were entered using the computer mouse and then registered in databases. The order of the tasks was counterbalanced between sessions across participants.

Sennheiser HD201 circumaural headphones were used for all participants. Prior to testing, the frequency response for the left channel of one set was tested using white noise generated with a Bruel & Kjaer Photon+ signal analyser sent to the headphones. The output from the headphones was captured with a Bruel & Kjaer 4153 artificial ear equipped with a Bruel & Kjaer 4192 half inch condenser microphone and a Bruel & Kjaer 2669 pre-amplifier. The output signal from the artificial ear was captured with the Photon+ and analysed with the RT Pro v7.2 real-time signal analysis package. The results of the analysis showed that the headphones presented a fairly flat frequency response, with no fluctuations of importance.

Before each experimental session, participants completed a short volume calibration procedure, in which a sequence of 6 Spanish words of varying length interspaced with 300 ms silences was presented. On hearing the sequence, participants had to indicate whether they were hearing the stimuli clearly and at a comfortable intensity. If this was not the case, the volume was adjusted until listeners reported that intensity was appropriately set. Once the volume had been adjusted, it was registered and fixed for the remainder of the experimental session. All perception experiments were set up and presented in OpenSesame, a cross-platform experiment builder that manages experimental packages written for Python (Mathôt, Schreij, & Theeuwes, 2012).

6.3.4. Procedures for phoneme monitoring

Condition: Segmental

Listeners were presented with VCV to VV continua (e.g., from [ˈu.ð̞o] to [ˈu.o]), for /b/, /d/ and /g/. They were told that they would hear sound sequences and that on each trial they had to decide whether a given consonant was present or not. They were also told to expect the consonants to sound as they would in an intervocalic context. Two buttons labelled “Sí” (yes) and “No” (no) were made available for each trial. Listeners completed a practice session with 5 tokens, presented in a randomized order, before each consonant block. The stimuli for the practice session included examples from the full length of the continuum. For the task, participants were presented with each 10 step continuum twice, with stimuli being presented in a randomized order, amounting to 60 trials in total (10 steps * 2 repetitions * 3 consonants). The order of the consonant blocks was counterbalanced across participants.

Condition: Word-level

For the most part, this condition was identical to the segmental condition. It differed in that, instead of VCV to VV sequences, listeners were presented with word-level continua (e.g., from *dudo* [ˈdu.ð̞o] to *dúo* [ˈdu.o]). Before the practice session, the complete list of target words was shown to participants, in order to prevent less frequent words (e.g., *cubetazo* or *boga*) from being affected by familiarity effects. Participants completed 60 trials (10 steps * 2 repetitions * 3 consonants).

Condition: Primed word

This condition was similar to the word-level condition, but it differed in that semantic primes were presented 300 ms before each word (see Figure 6.5), half of the time in favour of the full approximant interpretation (e.g., the prime *titubear* for *dudo*) and half of the time in favour of the elided interpretation (e.g., the prime *pareja* for *dúo*). Listeners were told that they were going to hear two words in a sequence for each trial and that they had to monitor for a consonant in the second.

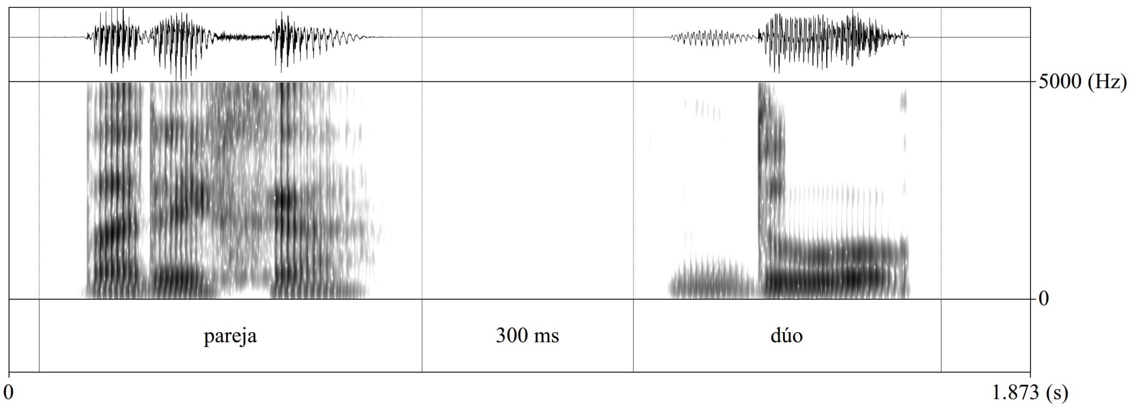

Figure 6.5. Semantic associate (prime) *pareja* ("couple") and target elided endpoint *dúo* ("duet") for phoneme monitoring, primed word condition. In this case, the prime favours the elided endpoint interpretation, and thus there is a match between the prime and the target.

A practice session with 10 tokens was completed for each consonant block, including priming for both ends of the continuum. The task consisted of a 10 step continuum for each consonant, presented two times for each prime type (full approximant priming and priming for elided variants) in a randomized order, amounting to 120 trials in total (10 steps * 2 repetitions * 2 primes * 3 consonants). Consonant blocks were counterbalanced across participants.

6.3.5. Procedures for identification

Condition: Segment

Stimuli, general presentation conditions and most instructions were the same as for the phoneme monitoring segmental condition. Listeners were told that they would hear VCV or VV sequences, and that they had to identify the sound. They gave their responses by clicking on two buttons containing an orthographic transcription for the two endpoints of the continuum (e.g., “ega”, for [ˈe.ɣ̞a], and “*ea*” for [ˈe.a]). Consonant blocks were counterbalanced across participants. A brief practice session with 5 tokens presented in a randomized order was completed before each consonant block. For the task, participants were presented with a 10 step continua twice, in a randomized order, totalling 60 trials (10 steps * 2 repetitions * 3 consonants).

Condition: Word-level

Stimuli and general presentation settings were the same as for the segmental condition, but with instructions and stimuli for a word-level condition. Listeners were asked to listen and identify a word from two options by clicking on buttons containing an orthographic transcription of a continuum endpoints (e.g., *mega* and *mea*). Each participant completed 60 experimental trials in a randomized order (10 steps * 2 repetitions * 3 consonants).

Condition: Primed word

Stimuli and general presentation conditions were the same as for the word-level condition, but semantic associates (primes) for both ends of each minimal pair (e.g., *titubear* for *dudo* and *pareja* for *dúo*) were presented 300 ms before each step. Listeners were told that they were going to hear two words in a sequence and that they had to identify the second word from two options transcribed orthographically on two buttons. Consonant blocks were counterbalanced across participants. For each consonant block, a practice session comprising 10 tokens was presented in a randomized order, with alternating priming for one or other end of the continuum. The task for each consonant consisted of a 10 step continuum presented twice for each type of prime, presented in a randomized order. Put another way, for two repetitions of each continuum the primes favoured one end of the continua and for the other two repetitions, the other end. In total, this task comprised 120 trials (10 steps * 2 repetitions * 2 primes * 3 consonants).

6.3.6. Procedures for discrimination

All discrimination tasks were designed under the ABX paradigm (Liberman, Harris, Hoffman, & Griffith, 1957; Creelman & Macmillan, 1979), which can be understood as consisting of a 2IFC subtask, in which the listener determines the order of the first two elements, and a yes-no subtask for the third item based on the decision made for the 2IFC subtask (Macmillan, Kaplan, & Creelman, 1977). The stimuli used for the ABX task were adapted from those used for phoneme monitoring and identification. Each 10 step continuum was transformed into 7 discrimination pairs with 2 step interval distances. The resulting pairs were 1-4, 2-5, 3-6, 4-7, 5-8, 6-9 and 7-10. A 300ms silence was inserted between the first two items, and a silence 700 ms long between the last two.

Condition: Segment

Listeners were instructed that they would hear 3-item-long sound sequences, interspaced with pauses, and that they had to determine whether the third element matched the first or the second. Participants were also instructed that the first item would always be different from the second, despite both being potentially very similar, and that the third one would always match either the first or the second. After each trial, participants saw two buttons to enter their responses: “Primero (A)” (first) and “Segundo (B)” (second). Consonant blocks were counterbalanced across participants. A brief practice session with 7 pairs was completed for each consonant block. The stimuli were presented in a randomized order, and all levels of the practice continuum were shown once. For the main task, participants were presented with 7 discrimination trials from the task continuum, also in a randomized order, including all permutations for the

ABX structure (ABA, BAA, ABB and BAB); this was done to control for bias towards the second element of the sequence, and to prevent primacy and recency effects (Postman & Phillips, 1965; Greene, 1986), as well as fatigue effects (Van der Linden, Frese, & Meijman, 2003). As a result, the task comprised 84 trials in total (7 pairs * 4 permutations * 3 consonants).

Condition: Word-level

For the most part, this condition was identical to the segmental condition, with the only difference that word sequences were presented to participants and instructions were adjusted accordingly.

Condition: Primed word

This condition mirrors the previous conditions in all respects, with the exception that a semantic prime was presented 300 ms before each ABX sequence (see Figure 6.6). Listeners were told to determine whether the fourth sound matched the second or third, and they were provided with buttons to enter their responses: “Segundo (A)” (second) and “Tercero (b)” (third).

Two types of preceding primes were presented: those favouring a full approximant interpretation and those favouring an elided interpretation. Half of the participants were shown the ABA and ABB permutations, and the other half the BAA and BAB permutations. Each participant completed 84 experimental trials (7 pairs * 2 permutations * 2 primes * 3 consonants).

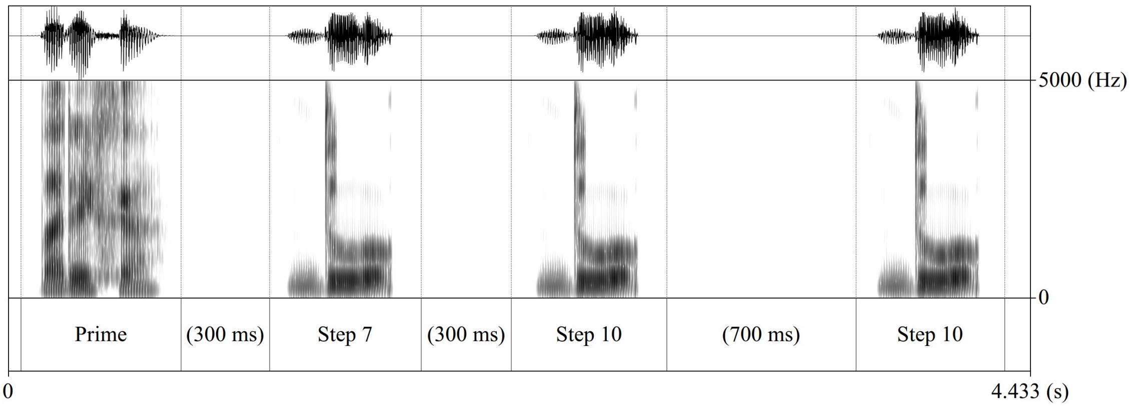

Figure 6.6. ABX trial for the discrimination task, primed condition. In this example, the semantic prime *pareja* is presented 300 ms before the ABX group, with trial stimuli from the continuum from *dudo* to *dúo*. In this case, steps 7 and 10 are presented first, and then step level 10 again after 700 ms, completing an ABB trial.

6.3.7. Summary of procedures

Ten-step continua for /b/, /d/ and /g/ from full approximant (e.g., *callado* [ka.ˈʝ̞a.ð̞o], “silent”) to elision from a minimal pair in which the elided variant from the first word rendered another (e.g., *Callao* [ka.ˈʝ̞a.o], “the Peruvian port”) were prepared for practice and experimental tasks, using acoustic models of natural speech and Klatt synthesis (lexical frequency was controlled within minimal pairs). Tasks and order of consonant presentation within session were counter-balanced across participants; stimuli within task and practice sessions were also randomized. Discrimination tasks were completed in one session, and phoneme monitoring and identification tasks in the other.

Stimuli were presented in four conditions: segmental, in which only the approximant consonant and neighbouring vowels were presented (e.g., a continuum from [ˈa.ð̞o] to [ˈa.o]); word level, in which the stimuli were embedded in a carrier word (e.g., a continuum from *callado* [ka.ˈʝ̞a.ð̞o] to *Callao* [ka.ˈʝ̞a.o]); semantic priming of the approximant consonant, in which a semantic associate primed the full approximant interpretation of the word-level continuum (e.g., *enmudecer*, “to silence”, priming *callado*, “silent”, in a continuum from *callado* to *Callao*); and semantic priming of the elided consonant, in which a semantic associate primed the elided interpretation of the word-level continuum (e.g., *puerto*, “port”, priming *Callao*, “the Peruvian port”, in a continuum from *callado* to *Callao*). Primes were presented 300 ms before the target; within each minimal pair, primes had a similar association strength to their targets.

Continua were presented in three tasks: phoneme monitoring, identification and discrimination. In the case of phoneme monitoring, listeners were asked to indicate whether they had perceived a target approximant consonant by the means of buttons labelled “yes” and “no” (10 step continuum * 2 repetitions * 4 conditions * 3 consonants = 240 experimental trials per participant). For identification, only the instructions and the response buttons were different: listeners were asked to select their response from two buttons each containing the labels for one end of the continua, transcribed orthographically (in total, 240 experimental trials were completed per participant). Finally, discrimination tasks were prepared under the ABX format (all permutations included), by converting each 10-step continuum into 7 discrimination pairs with 2 step interval distances. For segmental and word-level conditions, listeners were instructed to listen to the 3 items of each trial and to determine whether the third matched the first or the second (7 pairs * 4 permutations * 3 consonants * 2 tasks = 168 experimental trials per participant). In the case of the primed conditions, a semantic prime preceded the ABX trial, and listeners were asked to determine whether the fourth element matched the second or third; each participant was only presented with two ABX permutations, counter-balanced across participants (7 pairs * 2 ABX permutations * 2 primes * 3 consonants * 1 task * 2 conditions = 168 experimental trials per participant). A summary table listing details from all 72 task and practice continua is included in “Appendix 1: Full list of continua from Chapter 6” (see Table A1.1).

6.4. Results

As mentioned at the end of section “5.4.1. Identification of natural stimuli”, all GLMM analyses were conducted following a common procedure involving the determination of a best-fitting model, obtaining the statistical significance of main effects, random effects and interactions, and conducting post-hoc analyses for pairwise comparisons. For subsequent analyses, only a summary will be provided.

6.4.1. Phoneme monitoring

Phoneme monitoring: /b/

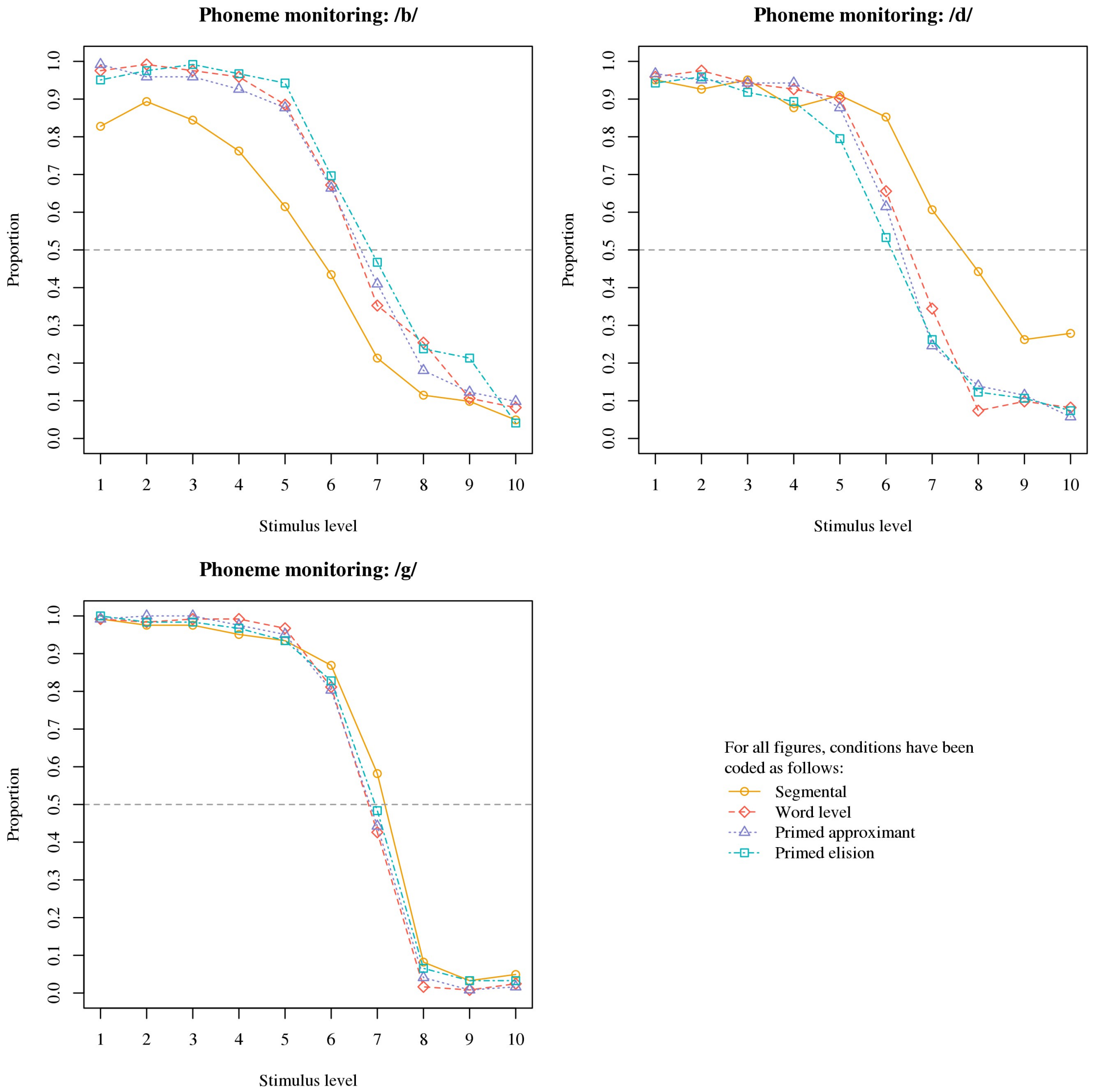

Figure 6.7. Phoneme monitoring results for /b/, /d/ and /g/, shown as averaged responses across participants (for each consonant, *n* = 4880). Proportion of reported presence of each consonant is shown as a function of stimulus level; chance level is shown as a dashed horizontal line.

The results for the [ˈu.β̞e] – [we] continuum in the segmental condition displayed a cumulative binomial distribution, with the first three steps around 85% perception of [β̞], lower than other conditions for which the starting point was around 95% percent perception (see top-left panel from Figure 6.7). Perception of [β̞] crossed the 50% chance level between steps 5 and 6, earlier than other conditions, and reached the lowest levels earlier, around stimulus step 8, with values close to 10%, and floor at step 10. Overall, perception of [β̞] in this condition was lower, even when the acoustic evidence for the full approximant was present in the signal in the first steps of the continuum.

For the word-level condition, the results also approximated a cumulative binomial distribution in which the first four steps of the continuum reached values around ceiling, crossed the 50% chance level of perception between steps 6 and 7, and gradually decreased to lower values to reach levels of around 10% perception on step 9. These results were very similar to the primed conditions. In the primed approximant condition, the prediction was that perception of [β̞] would be similar as at word-level, but that a category boundary shift would be present in favour of the full approximant interpretation. No such semantic priming effect was observed (see top-left panel from Figure 6.7). Instead, the results follow virtually the same distribution observed for word-level. Only in steps 1, 7, 9 and 10 was the perception of [β̞] higher for condition primed approximant, and only clearly for step 7.

The results for the primed elision condition did not display a clearly visible semantic priming effect in favour of elision. The prediction was that this priming would decrease the overall perception of [β̞]. Instead, the results were similar to those from the word-level and primed approximant conditions. Perception of [β̞] showed values around 95% until step 5 and crossed the 50% chance level close to step 7, to finally reach values close to floor responses for step 10. Interestingly, from step 3 onwards, the primed elision condition displayed higher values of [β̞] perception than word-level and primed approximant conditions, in a direction opposite to that anticipated; this is clearly visible in stimulus steps 5, 7 and 9.

A GLMM analysis was conducted on the results of phoneme monitoring from /b/. The best-fitting model included response as the dependent variable, experimental condition and stimulus level (as a continuous variable) as main effects, their interaction, subject as a random factor, and stimulus level and experimental condition as random slopes. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; no relevant deviations from normality were observed (excess kurtosis: 1.27, skewness: -0.24). The analysis revealed a significant main effect of condition ($\chi^2(3) = 76.350$, $p < 0.001$), stimulus level ($\chi^2(1) = 139.155$, $p < 0.001$) and a significant interaction between condition and stimulus level ($\chi^2(3) = 34.195$, $p < 0.001$). Wald *z* statistics were computed to obtain the statistical significance of differences in the response variable for the levels of experimental condition and their interaction with stimulus level (see Table 6.5).

Table 6.5. Wald *z* statistics for differences in the response variable in the phoneme monitoring task for /b/ for condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	3.283	0.490	6.704	< 0.001	***
	Segmental	Primed app.	3.127	0.490	6.380	< 0.001	***
	Segmental	Primed elision	3.653	0.523	6.985	< 0.001	***
	Word-level	Primed app.	-0.156	0.542	-0.287	= 2.322	
	Word-level	Primed elision	0.370	0.571	0.648	= 1.551	
	Primed app.	Primed elision	0.526	0.556	0.940	= 1.032	
Within interaction between condition and stimulus level	Segmental	Word-level	-0.305	0.069	-4.436	< 0.001	***
	Segmental	Primed app.	-0.288	0.068	-4.244	< 0.001	***
	Segmental	Primed elision	-0.310	0.070	-4.437	< 0.001	***
	Word-level	Primed app.	0.017	0.078	0.224	= 2.469	
	Word-level	Primed elision	-0.005	0.080	-0.066	= 2.841	
	Primed app.	Primed elision	-0.023	0.079	-0.287	= 2.322	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

The statistical analyses revealed that the segmental condition was significantly different from the word-level, primed approximant and primed elision conditions. No other significant differences were found between conditions. As to the interactions, the relationship between stimulus level and response was significantly different for the distributions of the segmental condition when compared to the word-level, primed approximant and primed elision conditions (no other interactions were significant). These results support the previously reported descriptives which singled out the segmental condition as different from the other conditions.

Phoneme monitoring: /d/

The results for the segmental condition resembled a cumulative binomial distribution (see top-right panel from Figure 6.7). Perception of [ð̞] started with values around 95% and remained close to this level for the first 6 stimuli, at which point perception decreased and crossed the 50% chance level between stimulus step 7 and 8. The lowest levels for the segmental condition were 30% of perception for the last two steps. In general, perception of [ð̞] was high and did not reach floor despite the acoustic evidence being very scarce or even absent in the last steps of the continuum.

In the word-level condition, the same continuum from the previous condition was presented, but embedded in a word-level context, from *dudo* [ˈdu.ð̞o] to *dúo* [ˈdu.o]. The results for the word-level condition displayed a cumulative binomial distribution. The maximum values of [ð̞] perception were reached in stimulus steps 1 to 5, at which point the perception of [ð̞] decreased abruptly, crossed the 50% chance level between steps 6 and 7, and reached the lower values and stabilized around 10% for the last three steps. When compared to the segmental condition, perception of [ð̞] in the word-level condition was lower for the second half of the continuum.

The results for the primed approximant condition, summarized in the top-right panel from Figure 6.7, show that [ð̞] perception reached its maximum levels in steps 1 to 4 with values around 95%. From this point onwards, perception decreased and crossed the 50% chance level close to step 6, and stabilized around 10% for the last three stimuli. Overall, responses from this condition were very similar to those from the word-level condition, with the exception of step number 7, where the responses for the primed approximant condition were lower, and step 8, where the inverse pattern was observed. In this condition, it was expected that the category boundary would shift towards perception of [ð̞] when compared to condition word-level, but this was not completely substantiated by the observed responses.

The responses for condition primed elision displayed a cumulative binomial distribution. The first four steps showed values around 90%, and then decreased gradually from step 5 onwards. Responses crossed the 50% chance level around step 6, and stabilized with low values around 10% for the last three steps of the continuum.

Overall, responses followed a similar pattern as word-level and primed approximant conditions, but perception was lower than these two conditions in the group of stimuli immediately around the category boundary. In this case, the prediction of lower perception of [ð̞] with respect to the word-level condition was met.

A GLMM analysis was conducted on the results of phoneme monitoring for /d/. The best-fitting model for this analysis included response as the dependent variable, experimental condition and stimulus level as main effects, their interaction, subject as a random factor, and stimulus level as a random slope. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; no relevant deviations from normality were observed (excess kurtosis: 1.27, skewness: -0.38). A significant main effect of condition was found ($\chi^2(3) = 24.626$, $p < 0.001$), along with a significant main effect of stimulus level ($\chi^2(1) = 106.361$, $p < 0.001$) and a significant interaction between condition and stimulus level ($\chi^2(3) = 72.581$, $p < 0.001$). Wald *z* statistics exploring the differences in the response variable for condition levels and their interaction with stimulus level are provided in Table 6.6.

Table 6.6. Wald *z* statistics for differences in the response variable in the phoneme monitoring task for /d/ for condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	1.813	0.421	4.308	< 0.001	***
	Segmental	Primed app.	1.581	0.411	3.845	< 0.001	***
	Segmental	Primed elision	0.682	0.379	1.799	= 0.216	
	Word-level	Primed app.	-0.232	0.455	-0.510	= 1.830	
	Word-level	Primed elision	-1.131	0.428	-2.643	< 0.05	*
	Primed app.	Primed elision	-0.899	0.418	-2.148	< 0.1	.
Within interaction between condition and stimulus level	Segmental	Word-level	-0.438	0.063	-6.995	< 0.001	***
	Segmental	Primed app.	-0.419	0.062	-6.806	< 0.001	***
	Segmental	Primed elision	-0.313	0.057	-5.453	< 0.001	***
	Word-level	Primed app.	0.018	0.070	0.264	= 2.376	
	Word-level	Primed elision	0.124	0.066	1.872	= 0.183	
	Primed app.	Primed elision	0.106	0.066	1.617	= 0.318	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

The analyses showed that the segmental condition was only significantly different from word-level and primed approximant conditions. Word-level was also significantly different from the primed elision condition. As to the interactions, the relationship between stimulus level and response were significantly different in the segmental condition than in the word-level, primed approximant and primed approximant conditions. These results confirm the observations from the statistical descriptives; firstly, that the segmental condition differs from all other conditions and, secondly, that there is a semantic priming effect for the primed elision condition.

Phoneme monitoring: /g/

The results for the segmental condition for /g/ showed a cumulative binomial distribution with values that centred on around 95% for the first 5 steps and then descended abruptly to cross the 50% chance level between stimuli 7 and 8. Perception stabilized close to 10% perception for the last three steps (see bottom panel from Figure 6.7). When compared to all other conditions, the segmental condition displayed slightly higher values of [ɣ̞] perception for the steps around the category boundary shift.

The first five steps of the continuum for the word-level condition displayed the highest perception of [ɣ̞] with values around ceiling. From step 5 onwards, perception decreased, crossed the 50% chance level close to step 7 and stabilized with values close to floor for the last three stimuli. The differences between this and the segmental condition are that the first and last sections of the continuum reached values closer to ceiling and floor perception, and that the higher levels of [ɣ̞] perception were sustained longer than for segmental condition.

The results of both primed conditions did not differ in any noticeable way from word-level condition. Levels of [ɣ̞] perception reached values around 95% perception for the first 5 steps, decreased abruptly and crossed the 50% chance category boundary level on step 7 and stabilized around 5% in the last three steps. When these values were compared to the word-level condition, no clear semantic priming effect was observed.

A GLMM analysis was conducted on the results of phoneme monitoring from /g/. The best-fitting model included response as dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, subject as a random factor, and stimulus level as a random slope. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; no important deviations from normality were observed besides some positive kurtosis (excess kurtosis: 5.03, skewness: -0.32). The results from this analysis failed to reach significance for condition as a main effect ($\chi^2(3) = 1.1319$, $p = 0.77$), but a significant main effect of stimulus level was found ($\chi^2(1) = 160.1113$, $p < 0.001$), as well as a significant interaction between condition and stimulus level ($\chi^2(3) = 160.1113$, $p < 0.001$). Wald *z* statistics were computed to obtain the statistical significance of differences in the response variable for the different levels of experimental condition and their interaction with stimulus level (see Table 6.7).

Table 6.7. Wald *z* statistics for differences in the response variable in the phoneme monitoring task for /g/ for condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	0.133	0.221	0.600	= 1.640	
	Segmental	Primed app.	0.193	0.224	0.859	= 1.173	
	Segmental	Primed elision	-0.007	0.203	-0.034	= 2.919	
	Word-level	Primed app.	0.060	0.245	0.245	= 2.418	
	Word-level	Primed elision	-0.140	0.227	-0.616	= 1.614	
	Primed app.	Primed elision	-0.200	0.230	-0.868	= 1.155	
Within interaction between condition and stimulus level	Segmental	Word-level	-1.201	0.359	-3.348	< 0.01	**
	Segmental	Primed app.	-1.274	0.365	-3.487	< 0.001	***
	Segmental	Primed elision	-0.440	0.302	-1.458	= 0.435	
	Word-level	Primed app.	-0.074	0.429	-0.171	= 2.592	
	Word-level	Primed elision	0.761	0.377	2.015	= 0.132	
	Primed app.	Primed elision	0.834	0.384	2.172	< 0.1	.

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

The analyses revealed no significant differences between experimental conditions. With regard to the interactions, the results showed that the relationship between stimulus level and response was significantly different in the segmental condition than in the word-level and primed approximant conditions, and that the primed approximant condition was significantly different from the primed elision condition. While it can be argued that the segmental condition shows a later category boundary than the other conditions, and thus an interaction with the word-level and primed approximant conditions was found, the interaction between the primed conditions is less clear, although it may be due to the values observed in steps 7 to 9, where the primed elision condition aligns better with the segmental condition than with the primed approximant condition.

6.4.2. Identification

Identification: /b/

For the segmental condition, results showed that [β̞] identification did not start at ceiling, but was around 85% for the first 4 steps (see top-left panel from Figure 6.8). Values then decreased abruptly until the 50% chance threshold was crossed between stimuli 5 and 6. Identification of [β̞] continued to decrease until it reached step number 8, where values stabilized around 10%, again not at floor identification. When compared to other conditions, the segmental condition displays lower identification levels across the continuum.

In the word-level condition, the results displayed a cumulative binomial distribution where the first 4 steps centred around 95% [β̞] identification and then decreased gradually. The 50% chance level crossing was located in step 7 and reached a stable point for the last 3 steps with values around 15%. When compared to the segmental condition, higher levels of identification were attained in all steps, which indicated an overall perceptual benefit from word-level semantic context, in line with predictions.

Results for the primed approximant condition were similar to those seen for the word-level condition. The first 4 steps displayed values close to 95% identification and then decreased gradually until they crossed the 50% chance level around step 7. The last three stimuli showed the lowest values of identification, descending from around 20% to 10%. The prediction for this condition was that the category boundary for [β̞] identification would occur later than in the word-level condition. However, results did not seem to support this expectation.

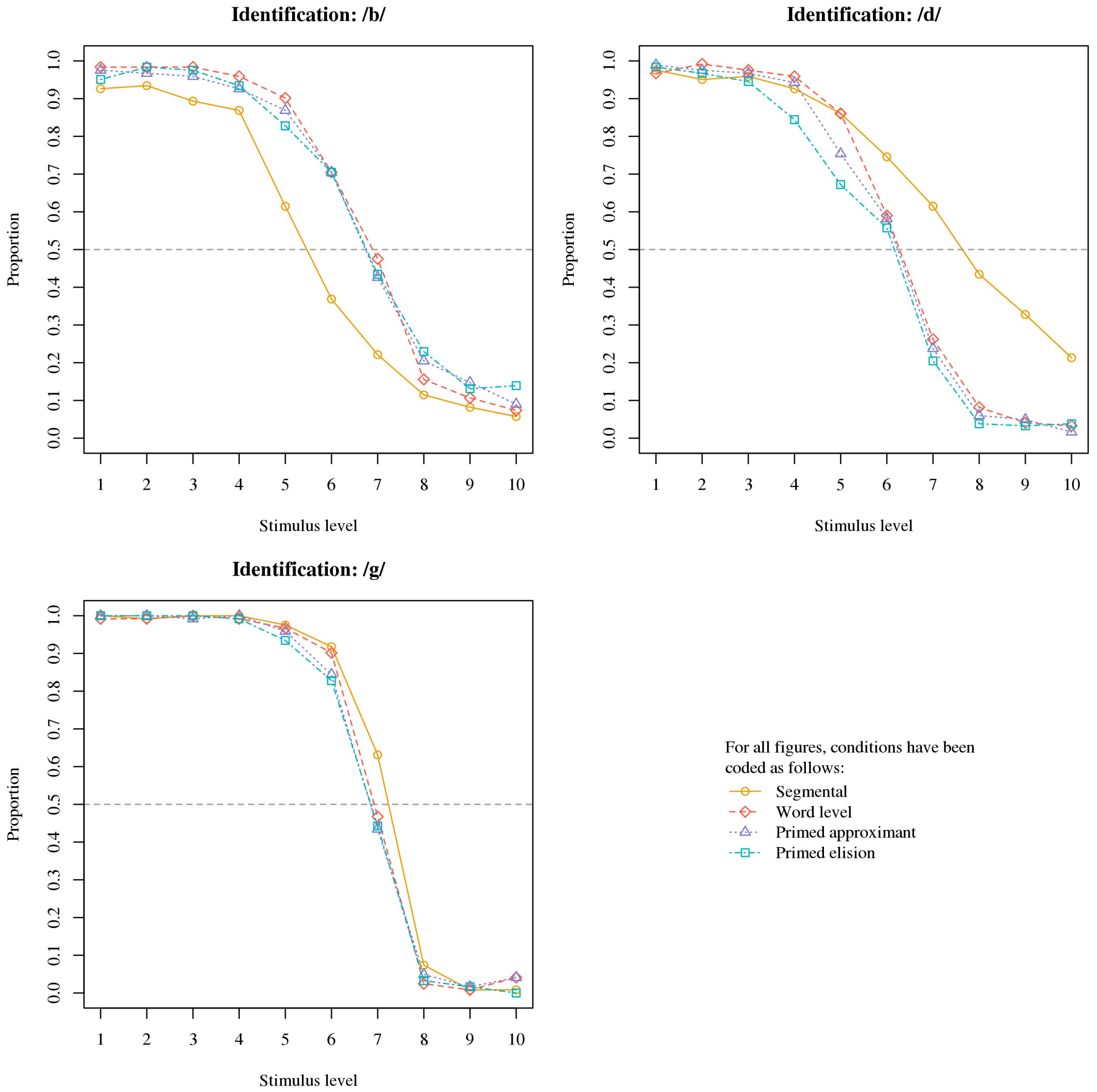

Figure 6.8. Identification results for /b/, /d/ and /g/, shown as averaged responses across participants (for each consonant, n = 4880). Proportion of identification for each consonant is shown as a function of stimulus level; chance level is shown as a dashed horizontal line.

The prediction for the primed elision condition was that the semantic prime *explosión* [eks.plo.ˈsjon], favouring the elided interpretation of the minimal pair, would shift the category boundary to earlier steps from the continuum with respect to the word-level condition. This shift was not observed in the results. Instead, the values for [β̞] identification showed a high degree of agreement with those from the primed approximant condition in most steps, with values around 95% identification for the first steps, a category boundary crossing at the 50% chance level in step 7, and final levels around 25% to 15% identification. Where the theoretical prediction does hold – step 5 and perhaps step 9 – the primed approximant condition displayed similar values.

A GLMM analysis was conducted on the results of identification from /b/. The best-fitting model included identification result as the dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, subject as a random factor, and stimulus level as a random slope. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; no relevant deviations from normality were observed (excess kurtosis: 1.38, skewness: -0.18). The results showed a significant main effect of condition ($\chi^2(3) = 41.867$, $p < 0.001$), a significant main effect of stimulus level ($\chi^2(1) = 205.054$, $p < 0.001$) and a significant interaction between condition and stimulus level ($\chi^2(3) = 14.649$, $p < 0.001$). Wald *z* statistics for differences in the response variable for the different levels of the independent variable experimental condition and its interaction with stimulus level are provided in Table 6.8.

Table 6.8. Wald *z* statistics for differences in the response variable in the identification task for /b/ for condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	2.874	0.468	6.139	< 0.001	***
	Segmental	Primed app.	1.545	0.402	3.847	< 0.001	***
	Segmental	Primed elision	1.210	0.388	3.120	< 0.01	**
	Word-level	Primed app.	-1.329	0.509	-2.609	< 0.05	*
	Word-level	Primed elision	-1.664	0.499	-3.334	< 0.01	**
	Primed app.	Primed elision	-0.335	0.437	-0.766	= 1.332	
Within interaction between condition and stimulus level	Segmental	Word-level	-0.250	0.071	-3.507	< 0.001	***
	Segmental	Primed app.	-0.059	0.063	-0.937	= 1.047	
	Segmental	Primed elision	-0.001	0.061	-0.017	= 2.961	
	Word-level	Primed app.	0.191	0.074	2.577	< 0.05	*
	Word-level	Primed elision	0.249	0.073	3.428	< 0.01	**
	Primed app.	Primed elision	0.058	0.064	0.900	= 1.104	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

The statistical analyses showed that the differences observed between the conditions were statistically significant, with the exception of the two primed conditions. As to the interactions, the results showed that in the word-level condition the relationship between stimulus level and response was significantly different from all other conditions. No other interaction was found to be significant. These results confirm previous observations. The interactions between the word-level condition and every other condition can be explained by the fact that its slope is, generally speaking, different from the slopes from the other conditions which seem to traverse in parallel. Instead, the word-level condition began aligned with the primed conditions, but ended closer to the values from the segmental condition.

Identification: /d/

The results for the segmental condition show that the first 4 steps reached around 95% of [ð̞] identification (see top-right panel from Figure 6.8). Values then decreased gradually and the category boundary crossing of the 50% chance level occurred close to step 8. The last two steps still displayed high values, with responses around 35% to 25% identification. Overall, [ð̞] identification was high when compared to other conditions, and the continuum did not reach a plateau close to zero.

In the word-level condition, results were organized in a cumulative binomial distribution. For the first 4 steps, [ð̞] identification approached ceiling, and then identification decreased gradually until it crosses the 50% chance level at step 6, and stabilized for the last 3 steps with values around 5% identification. From stimulus 5 onwards, identification values were lower than in the segmental condition.

The primed approximant condition, in which the semantic prime *titubear* [ti.tu.β̞e.ˈar] in favour of the approximant consonant was provided, displayed similar results to those from the word-level condition. The first four steps displayed values around 95% of [ð̞] identification. Then identification descended gradually until it crossed the 50% chance level around step 6, to finally settle around 5% identification in the last 3 steps. The prediction of higher identification for [ð̞] and a delayed 50% chance level category boundary crossing was not met by these results. For all steps, [ð̞] identification was virtually equal or lower than for the word-level condition, but higher than primed elision.

The results for primed elision condition also displayed a cumulative binomial distribution. These results were, in general, similar to those from the word-level and primed approximant conditions, with the first three steps showing values close to 95% identification and then descending gradually, to cross the 50% chance level close to step 6 and reaching stable values of floor identification for the last three steps. The prediction of less [ð̞] identification for this conditions was only met on steps 4 and 5, with the primed elision condition showing values lower than those from the word-level and primed approximant conditions.

A GLMM analysis was conducted on the results. The best-fitting model included identification results as dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, subject as a random factor, and stimulus level as a random slope. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; no relevant deviations from normality were observed (excess kurtosis: 2.01, skewness: -0.15). The

Table 6.9. Wald *z* statistics for differences in the response variable in the identification task for /d/ for condition levels and their interaction with stimulus level (SE = standard error).

Interaction	**Baseline**	**Comparison**	**Estimate**	**SE**	***z***	***p***	
No	Segmental	Word-level	2.711	0.495	5.477	< 0.001	***
	Segmental	Primed app.	2.352	0.442	5.320	< 0.001	***
	Segmental	Primed elision	1.191	0.398	2.990	< 0.01	**
	Word-level	Primed app.	-0.359	0.534	-0.673	= 1.503	
	Word-level	Primed elision	-1.520	0.499	-3.044	< 0.01	**
	Primed app.	Primed elision	-1.161	0.447	-2.598	< 0.05	*
Within interaction between condition and stimulus level	Segmental	Word-level	-0.620	0.075	-8.239	< 0.001	***
	Segmental	Primed app.	-0.610	0.068	-8.907	< 0.001	***
	Segmental	Primed elision	-0.474	0.062	-7.581	< 0.001	***
	Word-level	Primed app.	0.011	0.085	0.126	= 2.700	
	Word-level	Primed elision	0.147	0.080	1.828	= 0.204	
	Primed app.	Primed elision	0.136	0.074	1.844	= 0.195	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

results showed a significant main effect of condition ($\chi^2(3) = 43.093$, $p < 0.001$), a significant main effect of stimulus level ($\chi^2(1) = 148.523$, $p < 0.001$) and a significant interaction between condition and stimulus level ($\chi^2(3) = 123.816$, $p < 0.001$). Wald *z* statistics were used to explore the statistical significance of the differences observed in the response variable for different levels of the variable experimental condition and its interaction with stimulus level (see Table 6.9).

The results from the statistical analyses showed that all the differences observed between conditions were statistically significant, except for the comparison between word-level and primed approximant conditions. These results support the previous observations regarding the dissimilarities between the segmental condition and all other conditions, but also showed unexpected significant results for the other significant comparisons. As to the interactions, the results showed a significant interaction between the segmental condition and all other conditions, which is backed up by the delay in the perception of the elided variant.

Identification: /g/

For the segmental condition there was a cumulative binomial distribution curve, with values close to ceiling identification for the first 5 steps, and then an abrupt decline with a cross of the 50% chance level between steps 7 and 8, that stabilized with values of 10% to floor identification for the last three steps (see bottom panel from Figure 6.8). The remaining three conditions (word-level, primed approximant and primed elision) showed very similar results. The only differences were that, firstly, these three conditions crossed the 50% chance level slightly earlier than the segmental condition, and, secondly, that in both priming conditions [ɣ̞] identification decreased sooner than in segmental and word-level conditions (in steps 5 and 6).

A GLMM analysis was conducted on the results. The best-fitting model included identification results as dependent variable, experimental condition and stimulus level as main effects, subject as a random factor, and stimulus level as a random slope. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots; small deviations from normality were observed with some positive kurtosis (excess kurtosis: 7.64, skewness: -0.20). The results showed a significant main effect of condition on the identification results for /g/ ($\chi^2(3) = 18.642$, $p < 0.001$), and a significant main effect of stimulus level ($\chi^2(1) = 183.155$, $p < 0.001$). The results of Wald *z* statistics, calculated to explore the differences in the response variable for the different levels of the independent variable experimental condition, are provided in Table 6.10).

Table 6.10. Wald *z* statistics for differences in the response variable in the identification task for /g/ for condition levels and their interaction with stimulus level (SE = standard error).

Baseline condition	Comparison	Estimate	SE	*z*	*p*	
Segmental	Word-level	-0.495	0.192	-2.571	< 0.05	*
Segmental	Primed approximant	-0.604	0.193	-3.134	< 0.01	**
Segmental	Primed elision	-0.805	0.194	-4.152	< 0.001	***
Word-level	Primed approximant	-0.110	0.191	-0.573	= 1.701	
Word-level	Primed elision	-0.310	0.191	-1.620	= 0.315	
Primed approximant	Primed elision	-0.201	0.191	-1.050	= 0.882	

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

The results of the statistical analyses revealed that the segmental condition was different from the other conditions with statistical significance. No other differences were found to be significant.

6.4.3. Discrimination

Discrimination: /b/

In the segmental condition, the results showed low discrimination sensitivity for the stimuli across the stimulus pairs, with values around 60% discrimination (see top-left panel from Figure 6.9). Discrimination started closer to chance level values for pairs 1-4 and 2-5, and then rose to values closer to 60% and stabilized around that value for the remainder steps. The results for the word-level condition started with the first pair close to 60% discrimination and then decreased to chance level at the second pair, after which values increased gradually to around 65% for the last 4 pairs, surpassing discrimination sensitivity values found for the segmental condition. Overall, discrimination sensitivity increased when a full lexical context was provided.

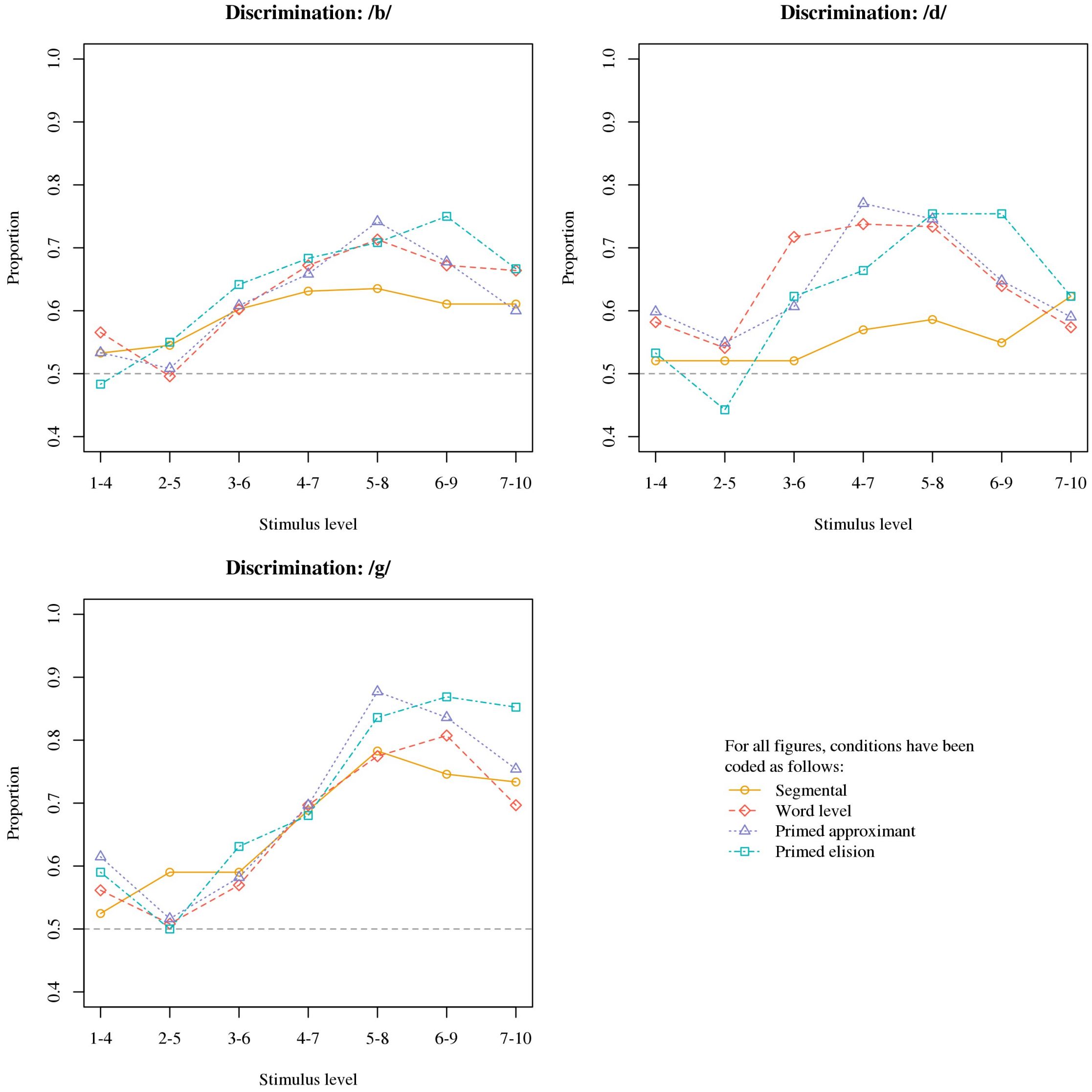

Figure 6.9. Discrimination results for variants of /b/, /d/ and /g/, shown as proportion of discrimination averaged across participants as a function of stimulus level pairs (for each consonant, *n* = 5124). Chance level is shown as a dashed horizontal line.

The results for the primed approximant condition showed a discrimination sensitivity at near chance level for the first 2 pairs and then a gradual increase until it reached a peak of 75% at pair 5-8 (fifth step). Sensitivity then decreased to values close to 60%. The discrimination sensitivity peak was more pronounced than in previous conditions. However, the expectations of a bias in favour of [β̞] perception, with a discrimination peak displaced to the left with respect to word-level condition, were not supported by the results. The primed elision condition showed values around chance level for the first pair, and increased to higher values until reaching a maximum of sensitivity around 75% for step 6-9. A decrease was then observed for the last pair, which showed values closer to 65% discrimination. Overall, discrimination values were higher than in the word-level and primed approximant conditions. For this condition, the prediction of a bias in favour of the elided interpretation of the minimal pairs was substantiated by the results, which showed a maximum discrimination peak closer to elision values when compared to the word-level condition.

A GLMM analysis was conducted on the results. The best-fitting model included discrimination results as dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, and subject as a random factor. The assumption of normality for the residuals was assessed via histograms and quantile-quantile plots. Serious deviations from normality were observed (see Figure 6.10), whose relevance will be discussed in section "6.4.4. Note on statistical analyses". The results of the analyses showed a significant main effect of stimulus level ($\chi^2(1) = 49.0668$, $p < 0.001$), but failed to reveal a significant main effect of experimental condition ($\chi^2(3) = 5.9166$, $p = 0.1157$) or a significant interaction between these two variables ($\chi^2(3) = 5.9380$, $p = 0.1147$). These results agree with the earlier observations regarding differences in the results for stimulus pairs, with discrimination increasing as a function of stimulus pair, but do not support the possible existence of differences between conditions.

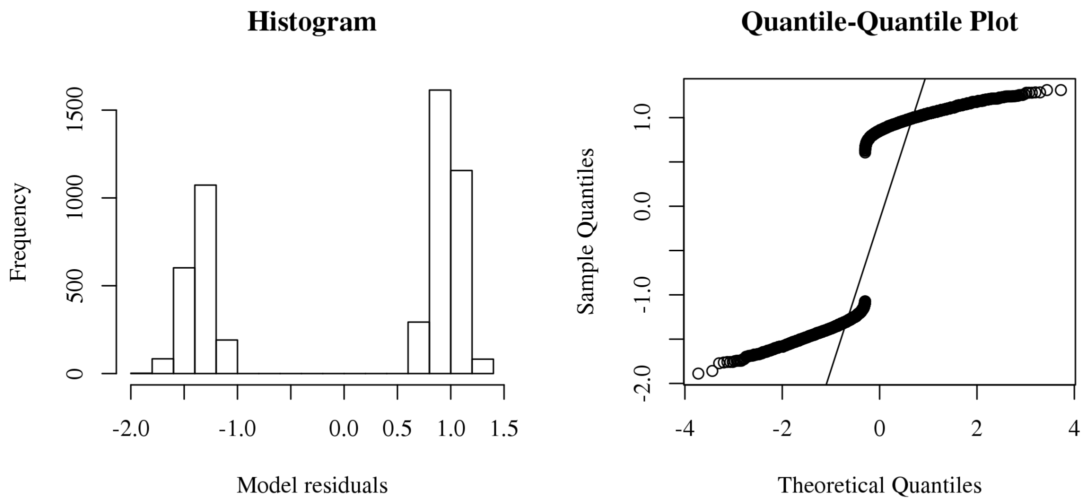

Figure 6.10. Histogram and quantile-quantile plot for the best-fitting model's residuals for the results from discrimination for variants of /b/ (n = 5097). The histogram shows a bimodal distribution for frequency values as a function of the model's residuals. The quantile-quantile plot shows the model's sample quartiles as a function of theoretical quantiles. A distribution close to a diagonal line in the direction shown is interpreted as fairly normally distributed.

Discrimination: /d/

The results from the segmental condition showed low discrimination sensitivity across the stimulus pairs, with values virtually at chance level for the first 3 pairs and then increasing to values around 60% discrimination, with no evident discrimination sensitivity peak (see top-right panel from Figure 6.9). In the word-level condition, discrimination began with values close to chance level for the first two pairs and then they increased to around 75% discrimination for pairs 3 to 5. In the last two pairs, discrimination decreased to values around 60%. The shape of this distribution might be indicative of a category boundary around the fourth pair.

The results for the primed approximant condition resembled those seen for the word-level condition, with the first pairs near chance discrimination, then an increase to values closer to 75%, and a fall towards 60% discrimination in the last two steps. The primed approximant condition showed a delay in the increase of discrimination sensitivity with respect to word-level condition (see step 3-6), and then it surpassed the maximum level of discrimination seen in this latter condition. The prediction of a bias in favour of the full interpretation of the stimuli with respect to the word-level condition was only partially backed up by the results. In the case of the primed elision condition, the first stimulus pair showed values close to chance level and then discrimination fell below this threshold for the second step. Afterwards, discrimination sensitivity rose gradually until reaching a discrimination peak in stimulus pairs 5-8 and 6-9, with values close to 75%. Finally, discrimination descended to a value closer to 60% in the last pair. The prediction of a bias towards the elided end of the continuum had some support from the results, with the discrimination peak shifted towards the right with respect to the word-level condition.

A GLMM analysis was conducted on the results of discrimination from /d/. The best-fitting model included discrimination results as dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, and subject as a random factor. Important deviations from normality were observed when the model's residuals were inspected using histograms and quantile-quantile plots (see Figure 6.11 and section “6.4.4. Note on statistical

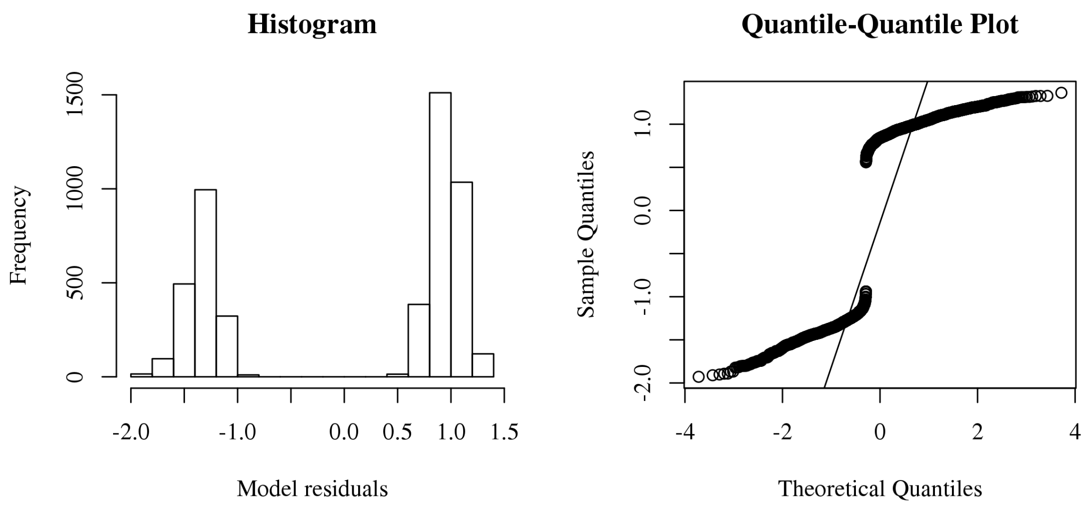

Figure 6.11. Histogram and quantile-quantile plot for the best-fitting model's residuals for the results from discrimination for variants of /d/ (n = 5124). The histogram shows a bimodal distribution for frequency values as a function of the model's residuals. The quantile-quantile plot shows the model's sample quartiles as a function of theoretical quantiles.

analyses”). The results from the analyses showed a significant main effect of stimulus level ($\chi^2(1) = 5.1921$, $p < 0.05$), a significant main effect of experimental condition ($\chi^2(3) = 16.1107$, $p < 0.001$) and a significant interaction between stimulus level and condition ($\chi^2(3) = 9.2174$, $p < 0.05$).

Wald *z* statistics were used to explore the statistical significance of the differences observed in the response variable for different levels of the variable experimental condition and its interaction with stimulus level (see Table 6.11). The results from the statistical analyses showed that the segmental condition was significantly different from the word-level condition, but not from any primed condition. Given the results observed in the top-right panel from Figure 6.9, the absence of a significant difference between the segmental and primed conditions is somewhat unexpected. As to the interactions, only the comparison between the word-level and primed elision conditions were statistically significant, most likely because of the sharp descent of the word-level condition results seen in pair 2-5, and a delay of the primed elision condition to reach lower sensitivity values in pair 6-9.

Table 6.11. Wald *z* statistics for differences in the response variable in the discrimination task for /d/ for experimental condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	0.514	0.158	3.249	< 0.01	**
	Segmental	Primed app.	0.426	0.194	2.194	< 0.1	.
	Segmental	Primed elision	-0.087	0.193	-0.452	= 1.956	
	Word-level	Primed app.	-0.088	0.196	-0.448	= 1.962	
	Word-level	Primed elision	-0.601	0.195	-3.087	< 0.01	**
	Primed app.	Primed elision	-0.513	0.225	-2.280	< 0.1	.
Within interaction between condition and stimulus level	Segmental	Word-level	-0.031	0.036	-0.867	= 1.158	
	Segmental	Primed app.	-0.011	0.044	-0.257	= 2.391	
	Segmental	Primed elision	0.102	0.044	2.305	< 0.1	.
	Word-level	Primed app.	0.020	0.045	0.443	= 1.974	
	Word-level	Primed elision	0.134	0.045	2.973	< 0.01	**
	Primed app.	Primed elision	0.114	0.052	2.197	< 0.1	.

Significance levels: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

Discrimination: /g/

The results for the segmental condition started at chance level for the first pair and increased gradually until reaching a discrimination sensitivity peak in pair 5-8 with around 80% discrimination (see bottom panel from Figure 6.9). Discrimination then decreased in the last two steps to values close to 75%. The results for the word-level condition showed discrimination starting slightly above chance level and then decreasing to chance level at the second pair. From pair 2-5 onwards, discrimination increased gradually until reaching a discrimination sensitivity peak around 80% in pair 6-9, after which it descended to around 70% discrimination. For the most part, responses showed a similar distribution than for the segmental condition.

For the primed approximant condition, discrimination began around 60% and then decreased to chance level for the second pair. From the third stimulus pair onwards, discrimination increased until it reached a maximum value in pair 5-8, approaching 90%, after which it decreased to levels closer to 75% discrimination. In line with predictions, the discrimination peak preceded that in the word-level condition. The results for the primed elision condition showed that the first stimulus pair had a discrimination value close to 60%. Discrimination decreased to chance level in the second pair and then it increased gradually until reaching a maximum around 85% for the last three steps. This maximum of sensitivity for the last pairs matches predictions of a category boundary shift in favour of an elided interpretation of the continuum.

A GLMM analysis was conducted on the results of discrimination from /g/. The best-fitting model included discrimination results as dependent variable, experimental condition and stimulus level as main effects, the interaction between condition and stimulus level, subject as a random factor and stimulus level as a random slope. Important deviations from normality for the model's residuals were observed in histograms and quantile-quantile plots (see Figure 6.12 and section "6.4.4. Note on statistical analyses"). The results from the analyses showed a significant main effect of stimulus level ($\chi^2(1) = 46.1917$, $p < 0.001$) and a significant interaction between stimulus level and condition ($\chi^2(3) = 8.8181$, $p < 0.05$), but failed to find a significant main effect of experimental condition ($\chi^2(3) = 2.0011$, $p = 0.57218$).

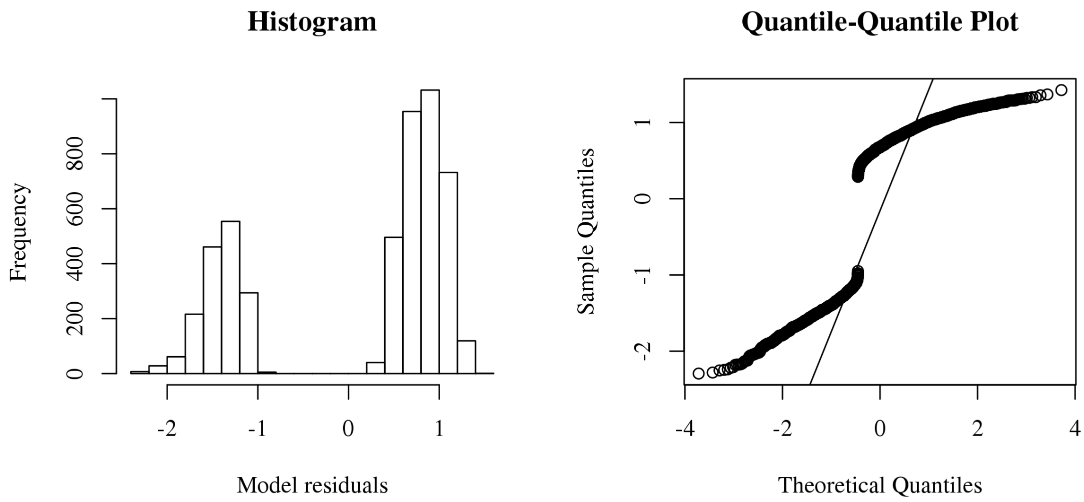

Figure 6.12. Histogram and quantile-quantile plot for the best-fitting model's residuals for the results from discrimination for variants of /g/ (n = 5124). The histogram shows a bimodal distribution for frequency values as a function of the model's residuals. The quantile-quantile plot shows the model's sample quartiles as a function of theoretical quantiles.

Wald *z* statistics were used to explore the statistical significance of the differences observed in the response variable for different levels of the variable experimental condition and its interaction with stimulus level (see Table 6.12). As expected given the results of the GLMM analysis, no significant differences were found between experimental conditions. As to the interactions, the relationship between stimulus level and response was significantly different in the primed elision condition when compared to segmental and word-level conditions. Differing trajectories for the first and last two pairs of the discrimination continuum can account for the significant interaction for the comparison between levels segmental and primed elision. The fact that the primed elision condition did not show a noticeable decrease in the last stimulus pairs is particularly important, as this also helps to explain the significant interaction when this condition is compared to the word-level condition.

Table 6.12. Wald *z* statistics for differences in the response variable in the discrimination task for /g/ for experimental condition levels and their interaction with stimulus level (SE = standard error).

Interaction	Baseline	Comparison	Estimate	SE	*z*	*p*	
No	Segmental	Word-level	-0.072	0.161	-0.450	= 1.959	
	Segmental	Primed app.	-0.043	0.199	-0.217	= 2.487	
	Segmental	Primed elision	-0.280	0.201	-1.392	= 0.492	
	Word-level	Primed app.	0.029	0.199	0.146	= 2.652	
	Word-level	Primed elision	-0.207	0.201	-1.033	= 0.906	
	Primed app.	Primed elision	-0.237	0.233	-1.020	= 0.921	
Within interaction between condition and stimulus level	Segmental	Word-level	0.012	0.038	0.311	= 2.268	
	Segmental	Primed app.	0.055	0.048	1.135	= 0.768	
	Segmental	Primed elision	0.139	0.050	2.791	< 0.05	*
	Word-level	Primed app.	0.043	0.048	0.888	= 1.122	
	Word-level	Primed elision	0.127	0.050	2.553	< 0.05	*
	Primed app.	Primed elision	0.084	0.058	1.462	= 0.420	

6.4.4. Note on statistical analyses

Generalized linear mixed model analyses –GLMMs– are not particularly new, but their computational availability and use within certain disciplines certainly is (Baayen, Davidson & Bates, 2008). As a consequence, there is still a lack of clear guidelines regarding several issues concerning their use, some fundamental, and some of lesser importance. Examples of core issues are discussions around best methods for model selection and whether it is possible to calculate degrees of freedom (Bolker et al., 2009), and also on tentative methods for calculating effect sizes (Nakagawa & Schielzeth, 2013). As to more superficial issues, there is considerable disagreement regarding how and what to report from GLMM results (for example, compare: Fucikova, Drent, Smits, & Van Oers, 2009; Poniatowski & Fartmann, 2009; Trisnawati, Tsukamoto, & Yasuda, 2015).

In theory, GLMM methods have only one assumption: that the residuals originating from the best-fitting model have to be approximately normally distributed (Quené & Van den Bergh, 2004). In previous research, two approaches have been followed regarding the assumption of normality for residuals: the approach in which the assumption is completely ignored when reporting GLMM implementations and results (e.g., Amo, Visser, & Oers, 2011; Berglund & Nyholm, 2011; Duffy, Cáceres, Hall, Tessier, & Ives, 2010; Fucikova et al., 2009; Natsumeda, Mori, & Yuma, 2012; Poniatowski & Fartmann, 2009; Poniatowski & Fartmann, 2011; Santos, Maia, & Macedo, 2009; Thünken, Meuthen, Bakker, & Kullmann, 2010; Trisnawati et al., 2015; Van Oers, Drent, Dingemanse, & Kempenaersm, 2008), and the minority approach in which they are reported, as in this thesis and rarely elsewhere (perhaps Ogura, 2012).

As mentioned in previous sections, the assumption of normality for the residuals was met for all phoneme monitoring and identification analyses, but not for those from discrimination, in which residuals described bimodal distributions (see Figure 6.10, Figure 6.11 and Figure 6.12). Non-normal residuals are to be interpreted, primarily, as a bad fit between the model and the data, which can be true for this data if we consider that, as explained in “5.4.1. Identification of natural stimuli”, all models were set for a *binomial* family, which means that the analyses try to fit a cumulative binomial distribution to the response distributions, in circumstances in which results from discrimination tasks differed considerably from said distribution (see Figure 6.9). Unfortunately, not many alternatives to GLMMs exist for our data from discrimination, considering that the data is non-parametric (binomial), and that both fixed and random effects need to be included into the models. For example, *d-prime* (*d'*), a well known bias-free signal detection sensitivity measure, often used with discrimination data (Macmillan et al., 1977; Stanislaw & Todorov, 1999; Macmillan & Creelman, 2005; Boley & Lester, 2009), cannot be calculated for each participant, which would have provided 61 data groups to analyse using alternative parametric tests such as repeated measures ANOVA. Using *d'* was not an option because each subject provided only two data points per discrimination pair, and consequently the proportion of their responses by pair could only take three values (0, 0.5 or 1), with which it is not possible to calculate reliable *hit* and *false alarm* rates. Another common alternative is using *arcsine transformation* for data from proportions (Sokal & Rohlf, 2003; for criticism see Warton & Hui, 2011), but again, transforming the data for each participant using this technique and then comparing it using parametric tests would provide meaningless results given the restricted values that the responses per discrimination pair display within participant. All in all, GLMM methods are likely the less harmful option for analysis of the discrimination data, but the statistical results need to be interpreted with caution. More responses by continuum step or discrimination pair will be collected in subsequent perception experiments, so as to be able to default to *d'* if the residuals from best-fitting models deviate from normality (see Chapter 7, but also Chapter 5, whose data was collected after the one reported here).

6.5. Summary of results

Phoneme monitoring

The overall results for phoneme monitoring showed that, in line with expectations, perception decreased as a function of continuum step. For most conditions, the 50% chance level was crossed around step 7, and then perception reached a plateau of minimum or no perception in the last steps of each continuum (see Figure 6.13). Although some differences for the results of /b/, /d/ and /g/ were observable, the

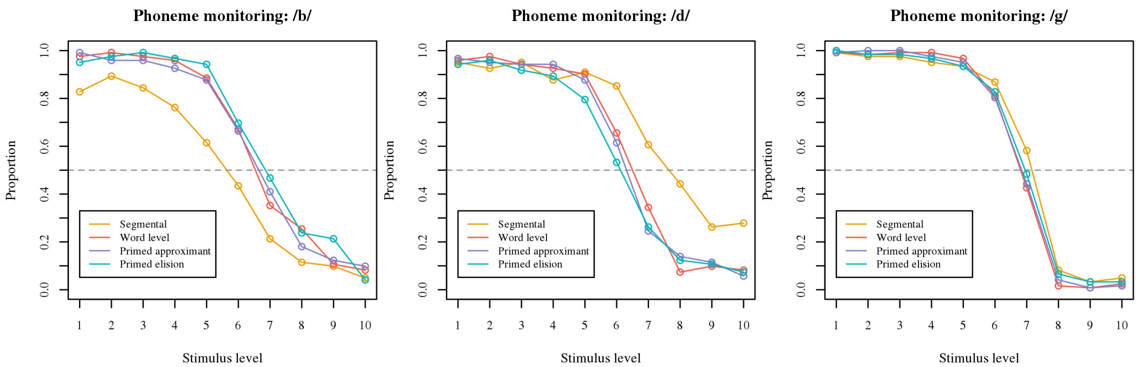

Figure 6.13. Summary of results for phoneme monitoring, showing means across participants for each consonant separately. Proportion of reported presence of the consonants is shown as a function of stimulus level, in continua from full approximant to elided variants, in four conditions: segmental, word-level, primed approximant and primed elision (n = 14640). Chance levels are shown as dashed horizontal lines.

word-level, primed approximant and primed elision conditions were remarkably similar for all consonants. The segmental conditions showed clear differences in the results for variants of /b/, /d/ and /g/. In the case of [β̞], perception in the segmental condition was lower, crossing chance level earlier. For the variants of /d/, the opposite effect was observed, with listeners showing signs of phonological recovery even for the last steps of the continuum. In the case of [ɣ̞], no clear differences between conditions were detected.

In line with descriptive results, GLMM analyses showed main effects of stimulus level in all consonants, and a main effect of condition for /b/ and /d/. A significant interaction between stimulus level and experimental condition was also found for the three consonants, surprisingly so in the case of /g/. Wald *z* post-hoc analyses revealed that the segmental condition was significantly different from most other conditions in the case of /b/ and /d/. In the case of /d/, the word-level condition was also significantly different from the primed elision condition, most likely due to differences observed between steps 3 and 8. No significant differences between conditions were observed for /g/, but some interactions were found, probably due to a delay in the perception of [ɣ̞] in the segmental condition and to small differences in the last steps of each continua.

Identification

The overall results for identification showed that proportion of identification decreased as a function of continuum step, with chance level crossings between steps 6 to 8, and final steps with minimum or floor perception (see Figure 6.14). Differences in discrimination were observed between phoneme category and between conditions within phonemic category. While word-level, primed approximant and primed elision conditions showed similar results for all consonants, the segmental condition differed in several respects. In the case of /b/, perception was generally lower than in the other conditions. For [ð̞], the opposite effect was observed, with listeners perceiving [ð̞] until later stimulus steps and finishing with values around 30% identification. In the case of [ɣ̞], no clear effect of condition was observable.

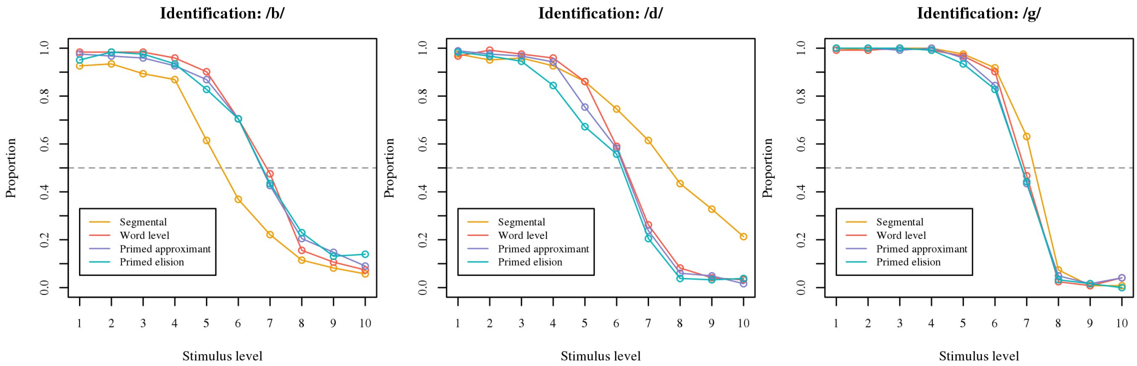

Figure 6.14. Summary of results for identification, showing results averaged across participants for each consonant and condition separately. Proportion of consonant presence is shown as a function of stimulus level, in continua from full approximant to elided variants, in four conditions: segmental, word-level, primed approximant and primed elision (*n* = 14640). Chance levels are shown as dashed horizontal lines.

The GLMM analyses revealed main effects of condition and stimulus level for all consonants. An interaction between these two variables was also found significant for /b/ and /d/, but not for /g/, in which all conditions displayed near identical trajectories across the continuum. Post-hoc Wald *z* analyses showed that the segmental condition was significantly different from all other conditions for the three consonants. The word-level condition was also significantly different from all other conditions for /b/. As to the interaction between stimulus level and experimental condition for /b/, significant differences in the variable response were observed for the comparison between the word-level condition and all other conditions. In the case of /d/, the segmental condition interacted with all other conditions.

Discrimination

Overall, results showed that discrimination sensitivity peaked as the amount of acoustic evidence available decreased (see Figure 6.15). These peaks were located in different step pairs, with a tendency for early peak discrimination sensitivity for /d/. Besides location, the magnitude of the peaks varied between consonants. Discrimination of /b/ variants showed lower values than discrimination for /g/, and performance was more similar within conditions. Discrimination for /d/ was somewhere in between, with large differences between conditions, some of which showed low discrimination sensitivity overall, as in the segmental condition, and others which showed varying degrees of discrimination sensitivity with clear discrimination peaks.

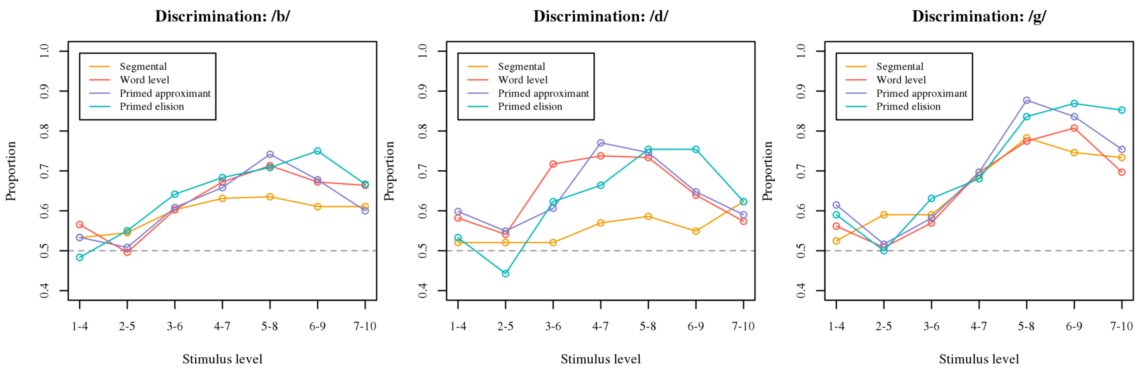

Figure 6.15. Summary of results for discrimination with mean proportions of discrimination across participants shown separately for each consonant and condition. Proportion of discrimination is shown as a function of stimulus level, in pairs of stimuli taken from continua from full approximant to elided variants (*n* = 15372). Chance levels are shown as horizontal dashed lines.

The statistical results from the GLMM analyses for discrimination showed a significant main effect of stimulus level for all consonants. A significant main effect of experimental condition was found for /d/, but not for /b/ or /g/, and a significant interaction between stimulus level and condition was found for /d/ and /g/. Post-hoc Wald *z* tests were conducted only for /d/ and /g/. In the case of /d/, the segmental condition was significantly different from the word-level condition only. The primed elision condition was significantly different from the word-level condition too. As to interactions, for /d/, the condition primed elision was significantly different from all other conditions. In the case of /g/, the primed elision condition was significantly different from segmental and word-level conditions.

6.6. Discussion

6.6.1. Interpreting the results

The perception experiments controlled a very specific set of cues in order to explore what is required for listeners to perceive approximant consonants of /b d g/. The condition with the smallest amount of available cues was the segmental condition, in which only the approximant consonant and surrounding segments were present. In this condition only the acoustic variables directly cueing for the approximant or its absence were available, along a minimal phonetic context which contained coarticulatory cues. This condition, particularly so in phoneme monitoring, provided an auditory baseline of perception, since no semantic or syntactic information was available to listeners to cue for the presence of approximants. Although comparing this condition to informationally more complex ones will be the focus of this discussion, on its own, it reveals important information about the differences that /b/, /d/ and /g/ display in perception.

Beginning with the results of the phoneme monitoring experiment for /g/ (see right-hand panel from Figure 6.13), listeners displayed a cumulative binomial distribution in which perception of [ɣ̞] started at ceiling, descended relatively fast until crossing chance level, and then further to floor perception for the last steps. So far, without considering other conditions, these results suggest that listeners take the acoustic evidence provided to them at face value, and that a continuum from presence to absence is perceived categorically (Liberman et al., 1957; Harnad, 1987). The hypothesis that the perception of /g/ is driven predominantly by acoustic cues gains support when results from identification and discrimination tasks in the segmental condition are also considered. In the case of identification, basically the same pattern of categorical perception was observed, besides slightly higher values in the early stages of the continuum and a small delay at chance level crossing in identification (see right-hand panel from Figure 6.14). In the case of discrimination, there was a coincidence between the location of a discrimination sensitivity peak for the segmental condition and the stimuli in which the chance level crossing occurred in phoneme monitoring and identification tasks, which is also consistent with a categorical perception hypothesis (see right hand panel from Figure 6.15).

Results are quite different for /b/ and /d/. In both, it is still true that acoustic cues have a strong effect on perception; the more acoustic cues are available for an approximant consonant, the more listeners report perceiving them, both in phoneme monitoring and identification (in both cases, this is backed up by main effects of stimulus level). However, acoustic information does not seem to be weighted as heavily as it did for /g/. The results of the segmental condition for /b/ in phoneme monitoring showed that listeners failed to perceive approximant consonants at ceiling when the acoustic information was fully available to them, and in general the perception of [β̞] was lower across the continuum, when compared to continua of /d/ and /g/ (see left-hand side panel from Figure 6.13). These results suggest that acoustic cues for [β̞] were less reliable when compared to those for [ɣ̞]. Results for the identification task for [β̞] in the segmental condition were similar to those from phoneme monitoring, except for one crucial difference: the first four steps of the identification continuum displayed a higher proportion of perception (see left-hand side panel from Figure 6.14). Since the stimuli presented in the segmental condition were identical for the two tasks, and only the instructions and type of response changed, these perceptual differences can be explained by task differences. While in phoneme monitoring listeners only had to indicate whether a segment had been presented, in the identification task listeners had to process orthographic labels from the response buttons prior to providing a perceptual judgement. A lexical effect on perception cannot explain an increase in perception, since both the VCV and VV sequences were Spanish nonsense words (Ganong, 1980). However, given their structure, the VCV sequences were more ecologically plausible than the VV sequences, which might explain the slight increase in perception of the consonant in the early stages of the continuum. It might be countered that the VCV sequence for [β̞] was indeed a word from Spanish (see the discussion around Table 6.4, in section “6.3.2. Stimuli”), but the same effect, albeit of lesser magnitude, was observed for /d/ and /g/ as well, where this was not the case.

Considering these results for variants of /b/ and /g/ in phoneme monitoring and identification, one might expect that the acoustic cues from [ð̞] would be even less reliable, since listeners rarely have full acoustic evidence for this consonant available to them (see Chapter 4), and therefore that perception of [ð̞] would display even lower values across the continuum when compared to [β̞]. However, results from the segmental condition displayed a pattern in the opposite direction: listeners perceived [ð̞] at near ceiling for at least half the phoneme monitoring steps, and then perception gradually descended, but never reached floor (see middle panel from Figure 6.13). In other words, [ð̞] was always perceived to some extent even when no acoustic cues for [ð̞] were available in the signal. Problems in the design of the stimuli –for example, a continuum biased in favour of the approximant end, failing to fully elide [ð̞]– could account for these results, but nothing from the acoustic characteristics of that particular continuum suggests that this was the case (see Table 6.2). Also, no traces of the word-initial [d] from [ˈdu.ð̞o] were present in the stimuli[18].

Two alternative explanations can be provided instead. First, it may be the case that listeners, knowing that evidence for /d/ is scarce in natural perception, are particularly sensitive to small acoustic cues for [ð̞], and thus require less evidence for perception by the end of the continuum. Second, it may be the case that listeners are not particularly sensitive to small acoustic cues for /d/, but that instead, knowing that evidence is scarce or unreliable, over-compensate for it after initial stages of speech processing, even when acoustic cues are completely absent from the signal (Mitterer & Ernestus, 2006; Janse et al., 2007). When the results from discrimination are considered, the second explanation seems more likely, because no evidence of particularly high sensitivity to small differences was observed in the results of discrimination for the segmental condition in [ð̞] (see middle panel from Figure 6.15). With the exception of slightly higher proportions of perception of [ð̞] in the early stages of the continuum from identification, that can be explained in the same way as for [β̞], no other differences of importance were observed between results of phoneme monitoring and identification.

Overall, providing a wider phonetic context and semantic cues in the word-level condition had the effect of bringing participants' perception of [β̞] and [ð̞] closer to distributions consistent with a categorical perception account, in both the phoneme monitoring and identification tasks. In the case of [β̞], the first section of the word-level continuum from phoneme monitoring reached ceiling perception, and then descended gradually to values close to floor (see left-hand panel from Figure 6.13). In the case of

18 In conditions where a word-level semantic context was available, participants were instructed to only pay attention to intervocalic approximants. If the word-initial [d] from *dudo* was actually priming perception of the approximant in these conditions, then higher proportions of perception would have been expected in word-level conditions and above. However, perception of /d/ decreased in comparison to the segmental condition.

[ð̞], the second section of the word-level continuum from phoneme monitoring crossed chance level sooner, and perception of the consonant reached floor in the last steps (see middle panel from Figure 6.13). Very similar results were observed for both consonants in the identification task, although the results of the phoneme monitoring experiments displayed more abrupt transitions from ceiling to decrease of perception and from decrease of perception to floor. Given that the only difference between the segmental and word-level conditions was the inclusion of additional phonetic and semantic contexts, it is safe to assume that the differences observed between those two conditions are due to these cues (Ernestus et al., 2002). In particular, the differences observed for [β̞] and [ð̞] when additional acoustic cues and a basic semantic context were provided can be explained as the result of two lexical representations, of similar lexical frequency, competing for perception in more or less equal terms, which results in two cancelling lexical effects on perception and a categorical treatment of the continuum. This would also explain why the results from the identification task are closer to a categorical perception distribution than in phoneme monitoring, given that a lexical effect should be stronger in a task where listeners are forced to process two lexical representations prior to providing their responses, as opposed to one in which listeners can ignore lexical representations.

Generally speaking, providing further semantic priming in two conditions in phoneme monitoring provided very little evidence of a priming effect in the hypothesized direction. In the case of [ɣ̞], the two primed conditions were nearly identical to the word-level condition, and thus it can be concluded that no semantic priming was present in the results (see right-hand panel from Figure 6.13). In the case of [β̞], the differences between the primed conditions and the word-level condition were, besides being small, in the opposite direction to the one expected (see left-hand panel from Figure 6.13). As for [ð̞], results showed some evidence of semantic priming for the primed elision condition, particularly between continuum steps from 3 to 6, but the primed approximant condition failed to show clear results in line with expectations, and stimuli towards the end of the distributions for both priming conditions displayed nearly identical results (see middle panel from Figure 6.13). Again, there was little evidence in favour of semantic priming in the identification results. The results from the semantic priming conditions of /g/ displayed slightly lower values than the word-level condition, but the two were virtually identical. For /b/, both priming conditions had slightly lower ceiling values and slightly higher floor values than those from the word-level condition (i.e., perception was less categorical). Except for a few continuum steps in the hypothesized direction (e.g., step 5), no clear evidence in favour of semantic priming was found (see left-hand panel from Figure 6.14). The results of /d/ showed lower perception values in the priming conditions when compared to the word-level condition, which is most clear in steps from 3 to 6 (see middle panel from Figure 6.14). For these steps, the relative position of the two priming conditions are consistent with a semantic priming hypothesis, which predicts lower perception of [ð̞] in the primed elision condition.

Adding semantic primes had the effect of lowering the perception of the three consonants relative to the word-level condition (although results from both phoneme monitoring and identification displayed this trend, it was clearer in identification). It has been argued above that adding word-level context made the two lexical items compete on more or less equal terms, and listeners started parsing continua categorically. Adding semantic primes on top of word-level cues degraded categorical perception to some extent, which may have resulted from a weaker combined lexical effect on (prelexical) speech processing of the two lexical items under competition, given the priming. Degraded as they may be, small effects of semantic priming were observed for /d/, in phoneme monitoring and identification, and were backed up by significant differences between the relevant conditions in phoneme monitoring, and by significant differences between the interactions of condition and stimulus level for relevant conditions in identification. As to the reasons why a small semantic priming effect was observed for /d/ and not the other consonants in these two tasks, one possibility is that semantic priming ought to be stronger for consonants for which the acoustic evidence is particularly unreliable in natural perception, as was the case for /d/. The opposite is true for /g/, with /b/ in a middle ground, with relatively reliable acoustic cues to perception of [β̞]. More powerful and direct priming alternatives such as embedding one end of the word-level continuum into a semantically congruent sentence in which the target becomes highly predictable might have shown stronger effects of semantic priming on speech perception (e.g., Warren, 1970; Warren & Sherman, 1974; Samuel, 1981a; Ernestus et al., 2002; Kemps et al., 2004).

As to the results of the discrimination tasks, in the case of /g/, sensitivity increased as a function of stimulus pair and reached a sensitivity maximum in those stimuli in which the chance level crossing occurred in phoneme monitoring and identification tasks (see right-hand panel from Figure 6.15). Increasing the amount of cues available at the word-level condition seems to have had the effect of slightly delaying and increasing sensitivity with respect to the segmental condition. Something similar happened in the priming conditions, which showed further increases in discrimination for those stimuli close to chance level crossings in the other tasks, and thus clear evidence of categorical perception (Harnad, 1987).

Discrimination results from /b/ were different. Discrimination peaks were less prominent than for [ɣ̞], although discrimination still increased as a function of stimulus pair, and discrimination maxima tended to coincide with chance level crossings observed in phoneme monitoring and identification (see left-hand panel from Figure 6.15). The segmental condition displayed low discrimination values, close to chance level along stimulus pairs, which suggests that listeners were not particularly sensitive to stimuli differences in this condition, while it also lends support to the idea that the acoustic variables directly cueing for the approximant consonant are relatively unreliable for /b/.

Overall, increasing the number of available cues in the word-level condition did have the effect of increasing sensitivity, particularly towards the second half of the continuum, as it did in the semantic priming conditions. Finally, the results for /d/ also displayed values close to chance level in the segmental condition, which could be interpreted as indicating that listeners were not sensitive to acoustic differences between variants of /d/ (see middle panel from Figure 6.15). Increasing the amount of cues for the word-level condition had a dramatic effect on discrimination. In particular, it increased discrimination sensitivity considerably for those stimuli involved in the chance level crossing in phoneme monitoring and identification. Semantic priming further increased discrimination sensitivity peaks with respect to the word-level condition, and shifted each distribution in line with predictions with respect to the last peak of the word-level condition. Taken together, results from discrimination showed four distinctive patterns: firstly, discrimination sensitivity increased for those stimulus pairs where the chance level crossing was observed in phoneme monitoring and identification; second, discrimination was lower in the segmental condition in all consonants, but near chance level for /b/ and /d/; thirdly, including a minimal full acoustic and semantic prime in the word-level condition increased discrimination evidence in favour of a categorical perception account; and finally, the two semantic priming conditions further increased discrimination, shifting discrimination peaks with respect to the word-level condition in the directions predicted by the primes.

The statistical results obtained from the GLMMs broadly support the interpretation of the results offered above. Main effects of stimulus level were found for all consonants and tasks, which lends credibility to the assumption that acoustic cues are a crucial contributor to the variability observed in the responses of the three tasks: overall, the more evidence there is for an approximant consonant, the more likely listeners are to perceive it, and the better participants are at discriminating between stimulus pairs. Main effects of condition were found for /b/ and /d/ in phoneme monitoring, as a result of differences between the segmental condition and most other conditions, and differences between the primed elision condition and the word-level condition for /d/ (all of which were significant in Wald *z* analyses). As a whole, the main effects and significant differences between conditions confirm that the segmental condition had a special status. No main effect of condition was found for /g/ in phoneme monitoring, given that all conditions display essentially the same distribution. In identification, a main effect of condition was found for all consonants, as expected for /b/ and /d/, but less so for /g/. Also, the segmental condition was significantly different from all other conditions for the three consonants, and the word-level condition was significantly different from all other conditions in /b/. For identification for /b/ and /d/, just as in phoneme monitoring, the statistical results support the observation that the segmental condition differs from all other conditions. For /g/, the main effect of condition can be explained by small differences between the segmental condition and the rest, which were larger than those seen in phoneme monitoring. In the case of the statistical results for discrimination (which should be considered cautiously), a main effect of condition was only found for /d/, consonant which displayed the largest differences between conditions, and the largest discrimination increase from the segmental condition to the rest. For this consonant, the segmental condition was significantly different from the word-level condition, and the primed elision condition was significantly different from the word-level condition.

Several interactions were also found in the results of the statistical analyses. In the case of phoneme monitoring, a significant interaction was found between stimulus level and condition for the three consonants. Post-hoc Wald *z* analyses revealed that these interactions were due to differences in the shape of the distribution of the segmental condition against all other conditions for /b/ and /d/, reinforcing the idea of a special status for this condition. For /g/, the interactions found in the post-hoc analyses revealed significant differences between the segmental condition against word-level and primed approximant conditions, due to a slightly late category boundary crossing in the case of the segmental condition. The interactions between stimulus pair and condition for the results of identification were only significant for /b/ and /d/, which was in line with expectations. In the case of /b/, this interaction was due to differences in the shape of the distributions between the word-level condition and all others, but surprisingly not between the distributions of the segmental condition and the two primed conditions. For /d/, the distribution in the segmental condition was significantly different from all other conditions, as was the case for the primed elision condition. The results of the interactions for the discrimination tasks are also in line with the observations previously made. Significant interactions between stimulus level and condition were found for /d/ and /g/, but not for /b/, where the distributions were relatively similar (no main effect of condition was found for /b/ either). In the case of /d/, the word-level condition was significantly different from the primed elision condition. For /g/, the distribution of the primed elision condition was significantly different from those in segmental and word-level conditions.

Summary

(a) When acoustic evidence in natural perception was reliable, the effect of adding contextual and semantic cues in experimental settings was practically null (see evidence for /g/).
(b) When acoustic evidence in natural perception was less reliable, perception was relatively lower in experimental settings. Additional contextual and semantic cues increased perception and phonological recovery (see evidence for /b/).

(c) When acoustic evidence in natural perception was particularly unreliable, listeners recovered phonological units for which the acoustic evidence was scarce or null in experimental settings. Adding additional contextual and semantic cues removed this effect (see evidence for /d/).
(d) Weak effects from semantic priming were only detected for a consonant with particularly unreliable acoustic evidence in natural perception (see evidence for /d/).
(e) Evidence for categorical perception increased when lexical effects from two comparable lexical competitors were present.
(f) Lexical effects were clearer in a task where post-lexical processing was mandatory (identification), and less clear in a task in which listeners could choose to ignore post-lexical processing (phoneme monitoring).
(g) Sensitivity to stimulus differences increased as the amount of acoustic evidence decreased in a continuum.
(h) Sensitivity to stimulus differences was generally low for consonants with unreliable acoustic cues in natural perception.
(i) Sensitivity to stimulus differences increased when semantic cues were provided.

6.6.2. Evidence for phonological recovery

Clear evidence of phonological recovery was found only for /b/ and /d/. In the case of /b/, it was observed in the results of phoneme monitoring and identification, particularly in the perceptual differences between the segmental condition and all others. Assuming that there is no reason to believe that listeners would display lower perception levels of [β̞] when it is fully cued acoustically (unless it was a bad exemplar of an approximant[19]), the segmental condition can be considered a baseline of how listeners perceive this segment when no additional cues are present. Increasing the cues available in the word-level condition, both in phoneme monitoring and identification, significantly increased the proportion of [β̞] perception across the continuum and shifted the chance level crossing until later steps (see left-hand panels from Figure 6.13 and Figure 6.14). This increase in perception constitutes a typical example of phonological

19 In any case, stimuli of poor quality has been shown to sometimes increase the strength of lexical effects on perception, which for /b/ would have meant higher perception (McQueen, 1996).

recovery (Warren, 1970; Samuel, 1981a). These results can be safely interpreted as phonological recovery, and not as a lexical effect favouring words instead of nonsense words (Ganong, 1980), because both lexical items were words, and their only difference was the presence or absence of /b/ in the underlying phonological representations. The way in which recovery manifested in the word-level and primed conditions differed in some details for phoneme monitoring and identification. In phoneme monitoring, in which no processing of the response labels was required, transitions from ceiling and floor to chance level crossing were more abrupt, probably given the auditory nature of that task (see left-hand panel from Figure 6.13). In identification, transitions were smoother, as a result of a stronger lexical effect from both competing lexical items in the perception results, which brought distributions from all conditions closer to what would be expected for categorical perception after recovery (see left-hand panel from Figure 6.14).

The case of /d/ is quite different. For this consonant, listeners reported more perception of [ð̞] than expected in the segmental condition. That is, they recovered an underlying unit for which there was no acoustic evidence in the signal. As argued earlier, this can be interpreted as an attempt by listeners to compensate for unreliable acoustic information cueing for [ð̞] in natural communication settings (see Chapter 3 and Chapter 4 for evidence from production). When a word-level semantic context was provided in the word-level condition, the perception of [ð̞] decreased with statistical significance for the second half of the continuum (i.e., phonological recovery receded), as the lexical effect of the elided end of the continuum started competing with that of the alternative lexical item, and distributions became categorical (see middle panels from Figure 6.13 and Figure 6.14).

6.6.3. Agreement with lexical access models

Episodic models

Episodic models assume that numerous exemplars are stored in long-term memory, and that similar episodes aggregate into clusters. These “memory clouds” are then matched to lexical representations for lexical access, after several processes have taken place. Episodic models deal with variation by assuming that exemplars for highly lenited forms are also stored in permanent memory. If these episodes are more frequent or recent, they will have a comparative advantage over alternative episodes (see "2.4.2. Episodic models" for more details). Under these models, it is expected that listeners have stored multiple exemplars of words containing /b/, /d/ and /g/, including episodes along the entire continua from approximants to elided variants. It is also expected that more exemplars exist for words containing the consonant's most frequent variants. For example, in the case of /g/ words containing open approximants should be better represented in exemplar clouds; in the case of /d/, absence of the consonant and very weak approximants should be better represented (see Chapter 4 for details).

When continua from approximant to elided variants are presented in isolation (as in the segmental condition), listeners cannot match the input to lexical-sized episodes. Instead, they have to match the segmental input to segmental-sized acoustic episodes, and better matches ought to occur for those steps from continua where stimuli are better represented by said episodes. Episodic models would thus predict the results observed in the segmental condition from /g/, and also the results observed for /d/ (since most episodes for this consonant contain little acoustic evidence, so not much is required to perceive [ð̞]), but not those found for /b/, where the full acoustic evidence cueing for [β̞] failed to reach ceiling perception and stimulus steps with less acoustic evidence (better represented in the episodic clouds) were not perceived as [β̞] as it happened for /d/.

In the case of word-level stimuli, listeners should find it easier to perceive inputs that agree with frequent episodes, although it is also expected that partial matchings to the input can occur, since there is more information available to make a lexical decision. These assumptions should apply more or less equally to both members of the minimal pair from each continua, given that they have similar lexical frequencies. They predict that listeners will perceive the continuum from /g/ categorically, since the acoustic cues for [ɣ̞] tend to be well represented in the signal in natural perception (when [ɣ̞] is present, acoustic evidence is clear). Instead, the category boundary for [β̞] and [ð̞] should be shifted to the right with respect to [ɣ̞], since variants with less acoustic evidence are good exemplars for both consonants (although harder to perceive). Only the predictions for /g/ were confirmed by the results. It is harder to evaluate the results from semantic priming, since mostly null effects were found, with the exception of /d/, which in any case were not backed up statistically. Also, the theoretical grouping of episodes under abstract lexical representations or labels is based on their acoustic similarity, and not on their semantic relatedness.

Strong episodic models of lexical access (e.g., LAFS) have trouble explaining lexical effects on speech perception, since labels representing episodic clouds do not intervene in prelexical stages of speech processing. Models where some top-down influence can take place, such as Minerva 2, are better prepared to account for lexical effects and phonological recovery by assuming that the echo of integrated exemplars that is returned after an episodic probe has been sent to long-term memory can contain information not present in the input episode, in essence, working as a pseudo-abstract representation. This semi-abstract echo would allow for the activation of lexical items for which there is imperfect matching. In our data, evidence for phonological recovery and lexical effects was observed for /b/ and /d/, and thus Minerva 2 is better suited to explain it, although not for the segmental condition, unless the probability of a sequence to conform a word could be evaluated against stored episodes. The influence of the two competing lexical items on perception, which made responses more categorical, was stronger in identification, a task in which evaluating the input against two underlying lexical representations was required before a response was provided. The fact that the two target lexical items were permanently activated in identification can explain a facilitatory effect in episodic models, in which the desired exemplar clouds would remain active throughout the task.

In episodic models, exemplars cluster naturally into groups defined by the similarity, frequency and recency of episodes. These groups can be thought to behave as prototypes, although they do not constitute abstract representations. If prototypes of some sort exist –albeit only functionally–, then episodic models should have no problem at accounting for categorical perception: listeners should be better able to tell apart examples from different episodic clouds than from the same clouds. In these results, sensitivity to stimuli differences increased as the amount of acoustic evidence decreased, in all conditions, but in some more clearly than in others. In the case of /g/, sensitivity maxima tended to coincide with identification category boundaries, which was interpreted as evidence for categorical perception. The results from the segmental condition for /b/ and /d/, in which proportion of discrimination was low overall, could be explained under episodic models by assuming that episodic clouds for the segmental level are not particularly strong or easy to match, since segmentation into intermediate abstract segmental categories is not an assumption of episodic models. However, it could be countered that LAFS does have separate spectral templates for alternative pronunciations and thus that low discrimination values should have been observed in the segmental condition for /g/ as well. When the acoustic evidence is provided in a word-level context, it becomes interpretable via matching to word-level sized episodes, and discrimination should improve, as was observed in the results.

Abstractionist models

Abstractionist models work under the assumption that the mental lexicon contains one abstract representation for each word, and that its structure consists of a string of abstract phonological segments. All models assume that the acoustic input has to be converted into a chain of some type of prelexical segmental-sized units that will later be compared to lexical abstract representations until an optimal match has been found and lexical access takes place. Beyond these commonalities, abstractionist models differ considerably in their posited processes and structures. For instance, most abstractionist models do not allow top-down feedback, i.e., Cohort, FLMP (Oden & Massaro, 1978; Massaro & Oden, 1980). Other models like RACE (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987) and Shortlist –both also autonomous– propose parallel and independent phonemic and lexical processing routes, whilst a minority implements top-down feedback directly (TRACE). Still others like Merge (Norris, McQueen, & Cutler, 2000) also have two parallel and independent processing routes like RACE, from prelexical processing units to lexical units, and from prelexical processing units to phoneme decision nodes, but also feedback from the lexical level to phoneme decision nodes (but not to prelexical acoustic processing nodes). Some abstractionist models such as Cohort and Shortlist make very specific predictions regarding how perception unfolds over time, allowing groups of candidates to compete depending on their degree of acoustic match to the incoming input. Lastly, different models defend alternative strategies to deal with lenited forms, ambiguous input, and to account for lexical effects and recovery. For instance, some models allow for underspecified features (Cohort), while others resort to independent lexical routes (RACE), facilitation from lexical levels to phoneme decision levels (Merge), or even to direct top-down feedback (TRACE).

Given that in these experiments all the sequences from the segmental condition were nonsense units, it is to be expected that only prelexical processing took place in their perception. Most abstractionist models of lexical access and speech perception are able to describe how listeners process segmental input by the means of prelexical processing modules that parse the acoustic input and extract the relevant features to build hypothesis of abstract phonological representations (the exceptions being Shortlist and Cohort). Models like RACE or Merge would predict categorical perception for the three consonants, since there is no clear way for an underlying phonological representation to contain degraded or partial features, which would otherwise explain the results from /d/, or a reason to expect clear instances of [β̞] to fail to achieve ceiling perception in the segmental condition. FLMP does accept partial featural and phonological matching, and thus it might be able to explain recovery for [ð̞] and the failure of [β̞] to reach ceiling in the segmental condition. In the case of /d/, because minimal evidence is still a better cue for [ð̞] than it is for its absence; in the case of [β̞], only if we were to assume that the full approximant was not a particularly good example of a [β̞]. Notice that inhibitory connections between competitors in feature and phonemic nodes, such as those modelled by TRACE, do not apply to the presence of an item competing with its absence, and thus they do not help explaining partial matching. In summary, abstractionist models of lexical access do not seem well suited to explain the results from segmental conditions, which is not surprising considering that most of the reviewed models are models of lexical access and not of speech perception.

The results from word-level conditions are better accounted for by lexical access models, although not all models are well prepared to predict the results from tasks in which lexical processing is not mandatory, as in phoneme monitoring. In Cohort, given that lexical access is achieved by the activation of all possible lexical candidates for an input, listeners cannot ignore lexical levels of processing in phoneme monitoring and provide a purely auditory response. Consequently, Cohort would predict identical results for the phoneme monitoring and identification tasks, prediction which was not supported by the results of the perception experiments described in this chapter. In identification, this model predicts that the two competing candidate words will be equally activated by the input (and primed by the response categories), and their activation will increase as a function of goodness of fit, until the disambiguation point, whereupon the acoustic evidence would be evaluated and a decision made, resulting in categorical perception, as observed in the data. Cohort is the only abstractionist model that refers explicitly to semantic priming. In particular, it predicts that semantic priming facilitates the activation of the primed candidate, even in the event of ambiguous input. The model would thus predict semantic priming effects for the three consonants, but they should be stronger for /d/, which was the only consonant in which semantic priming effects were clear.

In Shortlist, just as in Cohort, listeners cannot ignore lexical levels of processing, given that groups of lexical candidates activate as soon as the acoustic input is received, and thus this model is not very well suited to deal with pure auditory responses that do not require lexical access. In the case of identification, the model predicts that, as long as perception follows a prelexical route, the acoustic input will activate a group of candidates compatible with it, and that the competing lexical items will remain as good candidates until the acoustic input disambiguates in favour of one or the other, by the means of inhibitory links. Given that Shortlist is a RACE model, an independent lexical route should also be able to provide an output and thus account for lexical effects. In any case, Shortlist should predict categorical perception in identification.

Assuming that listeners parse the input of phoneme monitoring tasks resorting primarily to prelexical processing, Merge would predict that prelexical nodes will provide most information that phoneme decision nodes will receive (although feedback from lexical processing nodes to phoneme decision nodes cannot be ruled out), and that categorical response curves should be observed for a continuum from full approximant to elided variants, although categorical perception would not be maximized for phoneme monitoring because ambiguous input in this model is not resolved at early stages of speech processing. In identification, lexical decision nodes receive input from the prelexical processing nodes, and lexical nodes provide feedback to the phoneme decision nodes, disambiguating the input and producing responses closer to categorical perception, as observed in the data.

In the case of RACE, the model would predict that listeners will resort primarily to the prelexical processing route to provide a response in phoneme monitoring, but ambiguous input might allow the lexical processing route to achieve certainty thresholds first sometimes. Given that the two lexical items in competition are comparable in frequency and identical with the exception of the approximant consonant, it is still more likely for the prelexical route to win in phoneme monitoring, which would result in categorical perception distributions of responses. In identification, in which the two competitors are primed via the response categories, and given that sections of the continua are ambiguous, the lexical route should win more often, and provide results closer to categorical perception.

Finally, in the case of TRACE, in which information can flow in any direction between feature, phonemic and lexical nodes, the model would predict that in phoneme monitoring, in which responses can be purely prelexical, the input is processed by featural nodes and the presence or absence of the consonant could be resolved at the phonemic level. However, sometimes listeners may choose to adopt a post-lexical strategy in phoneme monitoring. In both cases, the model would predict categorical perception results in phoneme monitoring, but they should be clearer in identification, where there is a direct lexical effect in speech perception.

It is hard to evaluate whether abstractionist models can account for the discrimination results, mainly because these models concern themselves with the retrieval of features, phonological units and lexical candidates and not with how listeners process gradient details of the acoustic input. The issue of how sensitive are listeners to differences between similar sounds is reduced to whether a given acoustic signal can be mapped onto a phonological unit or not. In the case of continua from approximants to elided variants, it is to be expected that full acoustic evidence is mapped into underlying phonological units, and that this matching decays as acoustic evidence becomes scarce, until there is no sufficient acoustic evidence to extract the relevant features. In most models, categorical perception is expected (perhaps not in FLMP), and discrimination should behave accordingly, maximizing sensitivity to differences between stimuli belonging to different “categories”. As discussed several times before, some response distributions do seem to show patterns of categorical perception (e.g., the data for /g/), but in several other places this is not necessarily the case (e.g., the data from segmental conditions, and the contrast between the results from phoneme monitoring and identification).

Hybrid models

Hybrid models propose that both episodes and abstract representations exist and interact in speech perception and lexical access. In some models like Goldinger's CLS and Pierrehumbert's ED model, abstract representations always play a role in lexical access, while in other models like POLYSP, the retrieval of abstract linguistic units is an optional by-product of lexical access. Both Goldinger's CLS and POLYSP can be considered connectionist models, given that top-down feedback is possible, and consequently are particularly well suited to account for lexical effects on speech perception and phonological recovery; instead, Pierrehumbert's ED model is autonomous, given that the flow of information only goes from lower to higher processing levels. These models have the advantage that while exemplars can account for evidence showing that listeners do utilize phonetic detail in lexical access processes, and for learning, abstract representations can account for speaker-normalization and categorical perception.

These models, in general, have no trouble explaining the results obtained in the perception experiments. To begin with, the results from the segmental conditions in phoneme monitoring can be explained as for episodic models (as a result of the nature and structure of the episodic clouds for each consonant, which include expectations regarding what is expectable in natural perception). Adding semantic cues in the word-level condition should provide better matches to lexical-sized episodes, and also allow for top-down feedback to inform speech perception if listeners choose to use a lexical route for their responses (in CLS and POLYSP, at least). This explains responses describing distributions closer to categorical perception in the word-level condition. Given that lexical processing is mandatory in identification, lexical effects and recovery should be stronger than in phoneme monitoring for all consonants, and results in clear categorical perception distribution. Semantic priming should have the effect of increasing the relative advantage of the primed candidate by pre-activating the relevant episodic cloud, particularly when the acoustic evidence is unreliable, as was observed as a trend for /d/. The results from discrimination can also be explained by hybrid models. Firstly, increasing the amount of acoustic and semantic cues should facilitate discrimination, partly, because lexical-sized episodes can be compared. Consonants with relatively poor acoustic evidence in natural perception (/b/ and /d/) showed low discrimination values in the segmental level, which is expected for episodic clouds with relatively poor acoustic detail, which should be more difficult to discriminate.

6.6.4. Some limitations

As observed in “6.3.2. Stimuli”, homogenizing lexical frequency within minimal pairs greatly limited the number of candidate minimal pairs available, which was already low considering other restrictions that the stimuli had to meet (minimal pairs had to be members from different lemmas, approximants had to be located intervocalically, and the VCV and VV sections extracted from the minimal pairs had to not constitute legal words in Chilean Spanish). As a consequence, several minimal pairs had very low lexical frequencies (see Table 6.4), which might have contributed to the failure of semantic priming conditions to show consistent priming patterns beyond /d/.

In order to maximize semantic priming, an alternative design could be used, in which a strong semantic prime is presented before a high frequency lexical item, in a continuum from approximant to elision, and where completely eliding the approximant consonant results in a nonsense word, and thus lexical effects should be present in perceptual results (Ganong, 1980). Comparing the same continuum in two conditions, one primed and the other not, should reveal semantic priming effects on top of lexical effects, if they do exist.

Chapter 7

Follow-up study: Semantic priming in the perception of /b d g/

7.1. Introduction

As previously discussed (see “Chapter 2” and “Chapter 6”), non-acoustic variables can affect speech processing and categorical perception. For example, the status of a word can bias the perception of an acoustic continuum from word to nonsense word in favour of the word (Ganong, 1980; Samuel, 1981a; Fox, 1984; Connine & Clifton, 1987; Burton & Blumstein, 1995; Pitt, 1995; Connine, 1990; McQueen, 1991; Pitt & Samuel, 1993; Samuel, 1996). The amount of available contextual cues can also have an strong effect on perception, particularly for highly lenited forms and elided units. In particular, when the signal is not entirely reliable, listeners may resort to alternative sources of information to restore lenited or missing units, in what has been termed phonological recovery (e.g., Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004; Janse, Nooteboom, & Quené, 2007). Finally, phonological restoration has also been shown to be conditioned by word-frequency, given results showing more recovery in high-frequency words, although these effects have been small (Samuel, 1981a).

In previous chapters, several perception experiments exploring lexical effects and phonological recovery were conducted and their results analysed under the assumptions and predictions of lexical access models and models of speech perception. In these experiments, continua from full approximant to elided variants, in minimal pairs where both ends of the continua were legal Spanish words, were presented to participants in three tasks and several informational conditions, which included separate semantic priming of both ends of the continua. Overall, including semantic primes failed to render detectable semantic priming effects in the results, with the exception of /d/, in which the effects were nonetheless small and not entirely backed-up by statistical results. It was theorized that one possible reason to explain the failure of semantic primes at generating clear effects might have been related to the relatively low lexical frequency of the target lexical items from the minimal pairs, which was itself a consequence of homogenizing lexical frequency in order to control for lexical effects.

In this chapter a follow-up study will be conducted to explore whether increasing the lexical frequency of target items and providing stronger semantic primes produces semantic priming effects for approximant consonants of /b d g/. Continua from full approximants to elided variants will be prepared for high-frequency words. For these continua, removing the approximant consonant will not transform the lexical item into a different one, in contrast to previous experiments reported in Chapter 6. Continua will be presented in word-level and primed word-level conditions, in phoneme monitoring tasks. Given that only one end of each continuum can be interpreted as a word, lexical effects are expected to take place for all conditions, increasing consonant perception (Ganong, 1980). Semantic priming, if present, should have the effect of increasing perception in favour of the approximant consonant, shifting the perceptual category boundary to the right of each continuum and/or maximizing the contrast between consonant presence and consonant absence in perception. Testing the presence of semantic priming effects on perception will allow the evaluation of some of the claims and predictions that lexical access models make regarding the influence of higher levels of perception on lower levels of speech processing.

Aims

(a) Determine the effect of semantic priming in the perception of continua from approximant consonant to elided variants for Spanish /b d g/.
(b) Interpret the results in the light of lexical access models, with particular attention to their treatment of phonological recovery and lexical effects on speech perception.

7.2. Methods

7.2.1. Participants

Thirty monolingual native Chilean Spanish speakers participated in the perception experiments (mean age 20.4 years; 22 females and 8 males). Participants were undergraduate students residing permanently in Santiago, Chile. They received an information sheet prior to the experimental session, gave consent and completed a short questionnaire. None of them reported having any cognitive, hearing, language or speech impairment. Subjects were compensated for their participation.

7.2.2. Stimuli

Target words containing the sequences /a.ba/, /a.da/ and /a.ga/ were identified in the Chilean Spanish lexical frequency list LIFCACH (Sadowsky & Martínez, 2012). Given that the number of suitable words containing /a.ga/ was relatively small, target words containing /o.go/ were also identified and pre-selected from the same frequency list. Three target words for /b/, four for /d/ and three for /g/ were selected from a previously selected list of items, all of which had high lexical frequency and also had good semantic associates. These items were submitted to an online word association task in which 41 adult monolingual native Chilean Spanish speakers read each target word and provided the first associated noun they thought of. The two target words with the clearest semantic associate, as judged by agreement between the participants, were selected for each consonant. These target words and associates are summarized in Table 7.1. All selected words and primes had a relative lexical frequency above a threshold of 0.9 words per million. This threshold ensured that all items were contained within the most frequent 7.5%. Although this threshold might seem low, it is difficult to state whether it is actually conservative or not. For example, a threshold of 35 words per million has been used to divide corpora into high an low relative frequency sections containing 50% of the items (e.g., Stemberger & MacWhinney, 1986). While the threshold selected here is considerably lower, it isolates a much smaller section of the high-frequency items.

Table 7.1. List of target words and semantic associates (primes) for task and practice sessions. Agreement for the selected associates between participants is listed in the last column.

Phoneme	Status	Target word	Prime	Agreement
/b/	Task	*caballo* [ka.ˈβ̞a.ʝ̞o], "horse"	*herradura*, "horseshoe"	78.1%
	Practice	*trabajo* [tɾa.ˈβ̞a.xo], "job"	*empleo*, "employment"	61.0%
/d/	Task	*Adán* [a.ˈð̞an], "Adam"	*Eva*, "Eve"	53.7%
	Practice	*patada* [pa.ˈta.ð̞a], "kick"	*puntapié*, "kick"	24.4%
/g/	Task	*agarre* [a.ˈɣ̞a.re], "grip"	*sujete*, "to hold"	22.0%
	Practice	*diálogo* [ˈdja.lo.ɣ̞o], "dialogue"	*conversación*, "conversation"	12.2%

The selected words contained /b/, /d/ or /g/ in an intervocalic context, which should facilitate elision in connected speech (see Chapter 3 and Chapter 4). Additionally, all words for the experimental tasks had the same flanking vowels and the same lexical stress pattern[20]. The elision of the consonants from the selected words does not result in a separate lexical item; for instance, eliding /b/ from *caballo* [ka.ˈβ̞a.ʝ̞o] results in [ka.ˈa.ʝ̞o], which can only be interpreted in Spanish as the same lexical item with an elided /b/. As intended, this differs from the minimal pairs used in previous perception experiments reported in Chapter 6, such as *mega* [ˈme.ɣ̞a] against *mea* [ˈme.a], where eliding the approximant consonant results in [ˈme.a], which can be interpreted as either *mega* with an elided /g/ or as *mea*, a different lexical item altogether.

Target words and primes were recorded by the author, a native Chilean Spanish speaker, with the same settings as those reported in section "5.3.2. Stimuli", from Chapter 5. Target words and primes were excised manually in Praat using visual cues from waveforms and spectrograms (Boersma & Weenink, 2013). The approximant consonants from the target words and their neighbouring vowels (e.g., [a.ˈβ̞a] from

20 These restrictions, required to control for several confounding variables, reduced to some extent the number of available candidates, particularly for /g/, which is a relatively infrequent phonological unit in Spanish (Pérez, 2003).

caballo) were manually segmented, aided by visual cues from the waveform and spectrogram and by auditory inspection of the signals (see Figure 7.1).

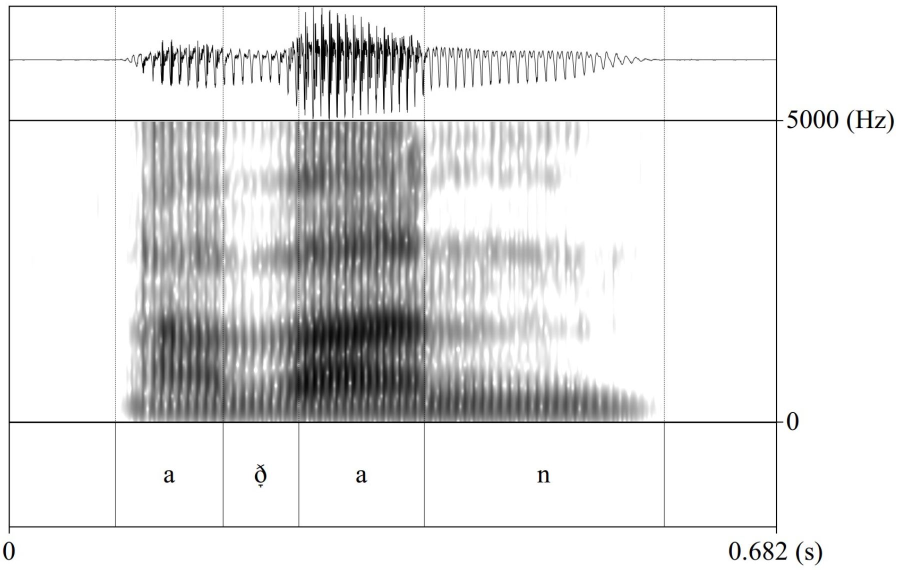

Figure 7.1. Manual segmentation for the word *Adan* [aˈð̞an], the full form target. The intervocalic approximant consonant [ð̞] is clearly visible in both the waveform and spectrogram as the intensity of the vocalic formants decreases from the second formant and upwards.

An acoustic model was built for each VCV section using 200 samples equally distributed along the temporal axis. For each sample, pitch, intensity, oral formants from F1 to F5, and oral formant bandwidths from F1 to F5 were measured. An acoustic model for an artificial elided variant was created by gradually merging the last third of the vowel preceding the approximant consonant and the first third from the vowel following it (see Figure 7.2). Each third was divided into 100 samples, where the same acoustic parameters as for the approximant consonants were measured.

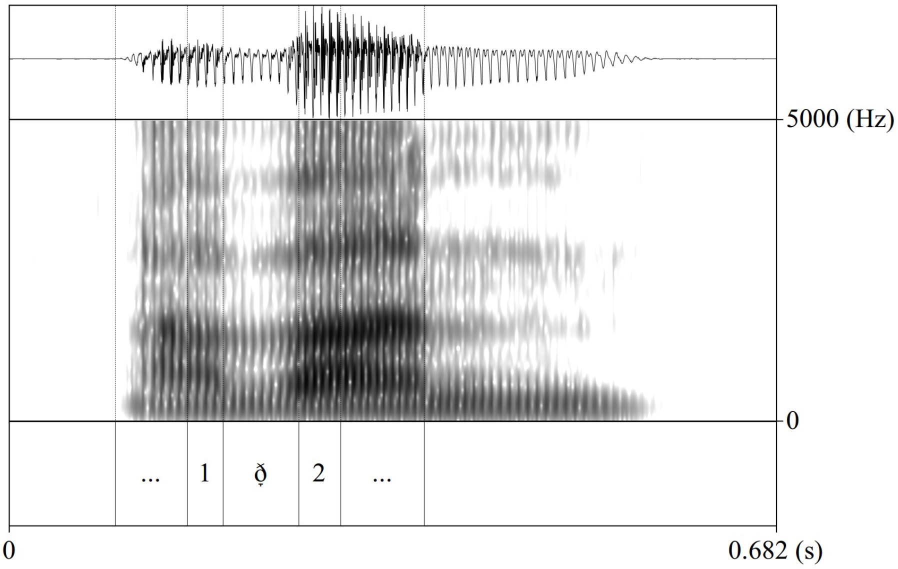

Figure 7.2. Waveform and spectrogram for the word *Adan* [aˈð̞an]. The last third from the vowel preceding the approximant consonant and the first third from the vowel following it have been segmented and labelled “1” and “2” respectively. The fully elided version for this stimulus was created by merging the two thirds so that it progressively shows less characteristics from the first section as the time domain moves forward, until it reaches the following vowel.

KlattGrid objects were populated for the approximant and elided models, and for 5 intermediate equally distributed steps. KlattGrids were converted to sound using Klatt synthesis (Klatt & Klatt, 1990; Weenink, 2009) and their intensity was scaled to that of the original VCV intensity. The resulting continuum steps were cross-spliced to the rest of the original word with an overlap of 10 ms. A short fade-out was included for stimuli in which the synthetic section was located at the final word boundary (as in *patada* and *diálogo*). Intensity was scaled to 70 dB and a Hann band-pass filter from 0 to 5000 Hz was applied to match the frequency range from the synthetic sections (see Figure 7.3).

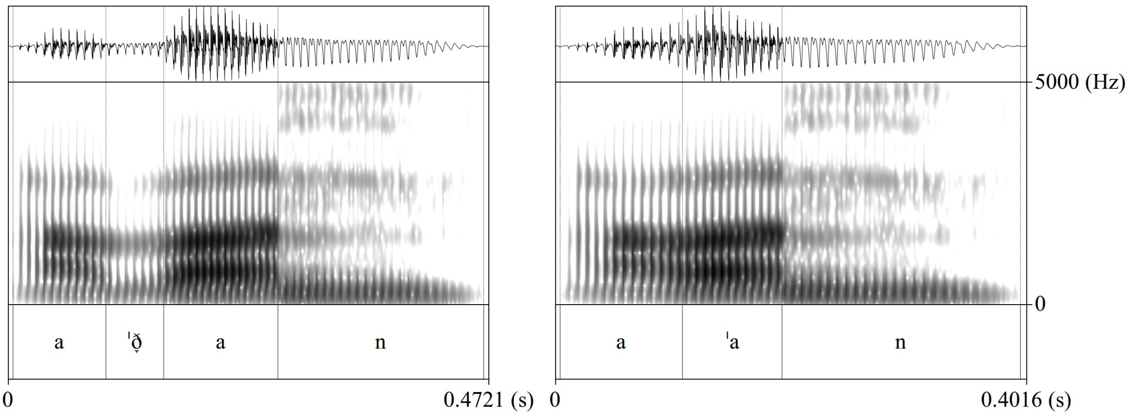

Figure 7.3. Waveforms and spectrograms for the synthesized and cross-spliced endpoints of the word to nonsense word continuum from [a.ˈð̞an] to [a.ˈan] for /d/. In the left-hand side panel, the approximant consonant is clearly visible as formant transitions and a general intensity decrease. On the right-hand side panel, the two neighbouring vowels meet and transition into each other without a noticeable intensity decrease. In both cases, the [n] segment has been spliced back with overlap.

For /b/, the relative lexical frequencies of the target word (*caballo*) and semantic prime (*herradura*) are 34.83 and 1.15, respectively (Sadowsky & Martínez, 2012). A summary of the main acoustic characteristics for the natural approximant consonant and the synthetic continuum from [a.ˈβ̞a] to [a.ˈa] can be found in Table 7.2. The first step of the synthetic continuum resembles natural **[β̞]** in duration, intensity, f_0, F1 and F2. The continuum steps gradually become more similar to the average of the neighbouring vowels, until full elision is attained. Oral formants from F3 upwards and bandwidths in step 1 differed from the natural stimuli, most likely for reasons similar to those provided in section "6.3.2. Stimuli": besides some of the stimuli being very short, the synthetic segments had simpler harmonic structures, which could have had an impact on the FFT analyses.

Table 7.2. Summary of acoustic characteristics for the approximant consonant **[β̞]** from [ka.ˈβ̞a.ʝ̞o], and synthetic **[β̞]** for each step of the continuum from [a.ˈβ̞a] to [a.ˈa], for the tasks involving /b/. Duration refers to the duration of the approximant consonant. Intensity to the minimum intensity within the consonant. Fundamental frequency, oral formant values from F1 to F5 and oral formant bandwidths from F1 to F5 are provided as means from the internal 50% duration of the consonant.

	[β̞]	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Duration (ms)	72.5	72.5	60.4	48.3	36.2	24.2	12.1	0
Intensity (dB)	56.5	63.3	65.1	66.6	68.1	69.2	70.3	NA
f_0 (Hz)	121.6	121.9	123.0	124.2	125.4	126.7	128.0	NA
F1 (Hz)	455	459	481	501	519	543	575	NA
F2 (Hz)	1170	1162	1192	1216	1235	1250	1271	NA
F3 (Hz)	2524	1264	1252	1265	1284	1299	1302	NA
F4 (Hz)	3844	2513	2507	2499	2499	2492	2476	NA
F5 (Hz)	NA	3778	3761	3733	3699	3672	3656	NA
$F1_{bw}$ (Hz)	167	138	133	128	126	125	127	NA
$F2_{bw}$ (Hz)	123	513	135	128	125	127	130	NA
$F3_{bw}$ (Hz)	408	1504	1895	1754	1457	1325	1466	NA
$F4_{bw}$ (Hz)	352	326	320	300	261	234	240	NA
$F5_{bw}$ (Hz)	NA	428	408	345	258	213	212	NA

In the case of /d/, the relative lexical frequencies of the target word (*Adán*) and the semantic prime (*Eva*) are 3 and 8.22 respectively (Sadowsky & Martínez, 2012). A summary of the main acoustic characteristics for the synthetic continuum from [a.ˈð̞a] to [a.ˈa] and for the natural [ð̞] can be found in Table 7.3. Duration decreased as a function of stimulus step, while intensity increased as the approximant became elided. The oral formant and bandwidths of the first synthetic step do not match the acoustic characteristics of the natural [ð̞]. This was most likely due to measurement differences in the availability of additional vocalic harmonics for the synthetic stimuli. Additional auditory analyses confirmed that the synthetic full approximant was a clear example of [ð̞].

Table 7.3. Summary of acoustic characteristics for the approximant consonant **[ð̞]** from [a.ˈð̞an], and for each step of the synthetic continuum [a.ˈð̞a] – [a.ˈa], for the tasks involving /d/.

	[ð̞]	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Duration (ms)	77.5	77.5	64.6	51.7	38.8	25.8	12.9	0
Intensity (dB)	55.7	63.3	66.2	67.4	68.3	68.8	69.2	NA
f_0 (Hz)	124.0	117.4	117.8	118.5	119.1	119.7	120.4	NA
F1 (Hz)	348	241	250	257	264	271	282	NA
F2 (Hz)	1342	1678	1691	1676	1660	1679	1714	NA
F3 (Hz)	2655	2704	2708	2722	2726	2724	2747	NA
F4 (Hz)	3899	4243	4197	4100	4107	4138	4157	NA
F5 (Hz)	NA	NA	NA	NA	NA	NA	NA	NA
$F1_{bw}$ (Hz)	248	69	81	95	107	118	132	NA
$F2_{bw}$ (Hz)	82	567	636	608	522	445	357	NA
$F3_{bw}$ (Hz)	199	142	130	129	122	103	70	NA
$F4_{bw}$ (Hz)	363	821	684	471	328	377	399	NA
$F5_{bw}$ (Hz)	NA	NA	NA	NA	NA	NA	NA	NA

For /g/, the relative lexical frequency of the target word (*agarre*) is 0.93. The semantic prime *sujete*, with that inflection in particular, was not present on the lexical frequency list. However, the non-inflected parent lemma *sujetar* has a lexical frequency of 15.6 words per million. A summary of the main acoustic characteristics for the synthetic continuum from [a.ˈɣ̞a] to [a.ˈa] and for the natural [ɣ̞] can be found in Table 7.4. As for the previous continua, duration decreased and intensity increased as a function of stimulus step. For the most part, the first synthetic stimuli matched the natural [ɣ̞]. However, some noticeable differences were observed for F1, and more clearly for all bandwidths.

Table 7.4. Summary of acoustic characteristics for the approximant consonant **[ɣ̞]** from the natural [a.ˈɣ̞a.re] recording, and **[ɣ̞]** for each step from the synthetic continuum between [a.ˈɣ̞a] and [a.ˈa], for the tasks involving /g/.

	[ɣ̞]	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Duration (ms)	62.7	62.7	52.2	41.8	31.3	20.9	10.4	0
Intensity (dB)	51.3	69.5	70.4	71.8	71.5	71.8	72.8	NA
f_0 (Hz)	115.5	119.9	120.5	121.2	122.0	123.2	125.1	NA
F1 (Hz)	366	460	465	469	473	476	479	NA
F2 (Hz)	1702	1712	1684	1653	1621	1583	1550	NA
F3 (Hz)	2342	2448	2457	2471	2483	2493	2503	NA
F4 (Hz)	4049	3787	3650	3507	3276	2879	2826	NA
F5 (Hz)	4421	4049	4068	4135	4120	4147	4149	NA
$F1_{bw}$ (Hz)	135	49	48	49	50	52	64	NA
$F2_{bw}$ (Hz)	392	100	85	81	82	76	72	NA
$F3_{bw}$ (Hz)	906	243	209	183	168	172	176	NA
$F4_{bw}$ (Hz)	1699	435	1317	1159	1535	2147	2062	NA
$F5_{bw}$ (Hz)	NA	NA	NA	NA	1742	344	387	NA

7.2.3. Procedures

The experimental sessions were all conducted in the Phonetics Laboratory of *Pontificia Universidad Católica*, in Santiago, Chile, by a trained phonetician (not the author). The general procedures replicate those from the experiments reported in the previous two chapters. Before each experimental session, participants completed a volume calibration procedure, in which a sequence of words interspaced with 300 ms silences was presented, until they reported hearing it clearly and comfortably. These words had been recorded by the author and their intensity normalized to 70 dB. All perception experiments were set up and presented in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012).

In the word-level condition, participants were instructed to listen to sound sequences and to monitor for a particular consonant (either /b/, /d/ or /g/). Instructions were given to participants to expect the consonants to sound as they would in an intervocalic context, and to only pay attention to intervocalic consonants. Participants completed a short practice session with a randomized 7 step continuum before each consonant block. The task continuum was presented 15 times in a randomized order, totalling 315 trials for this condition (7 steps * 15 repetitions * 3 consonants). To enter their responses, participants were provided with 2 buttons labelled “Sí” (yes) and “No” (no). Consonant blocks were counter-balanced across participants. The primed word-level condition was identical in all respects to the word-level condition, with the exception that a semantic prime was shown 300 ms before each target word (see Figure 7.4).

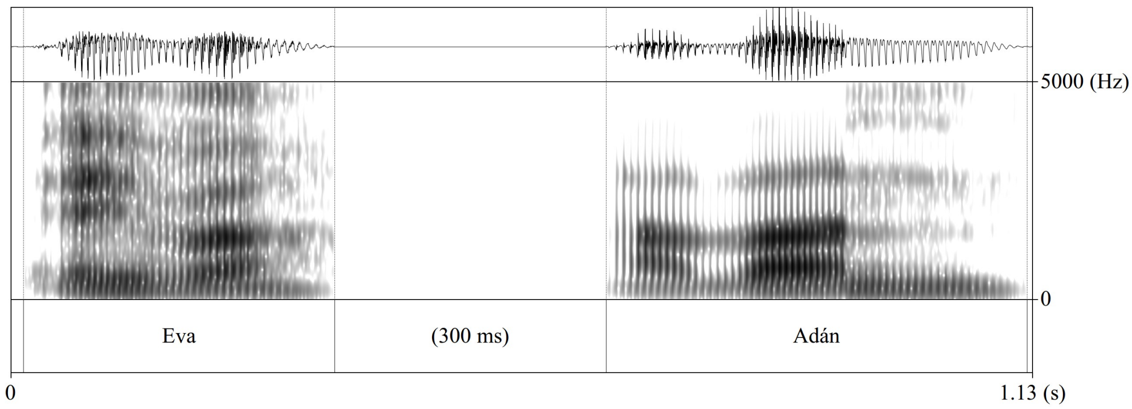

Figure 7.4. Example of a phoneme monitoring trial for the primed word-level condition. In this case, the semantic prime *Eva* was presented 300 ms before the first step from the continuum from [a.ˈð̞an] to [a.ˈan].

7.3. Results

Results for /b/

The results for the word-level condition differed from a cumulative binomial distribution. The first four steps of the continuum reached values around ceiling, crossed the 50% chance level between steps 6 and 7, to finally settle at around 30% perception on step 7 (see top-left panel from Figure 7.5). The primed word condition essentially replicated the results of the word-level condition, although slightly higher values of [β̞] perception were observed for the first continuum steps. The prediction of a category boundary shift in favour of perception of [β̞] was not observed.

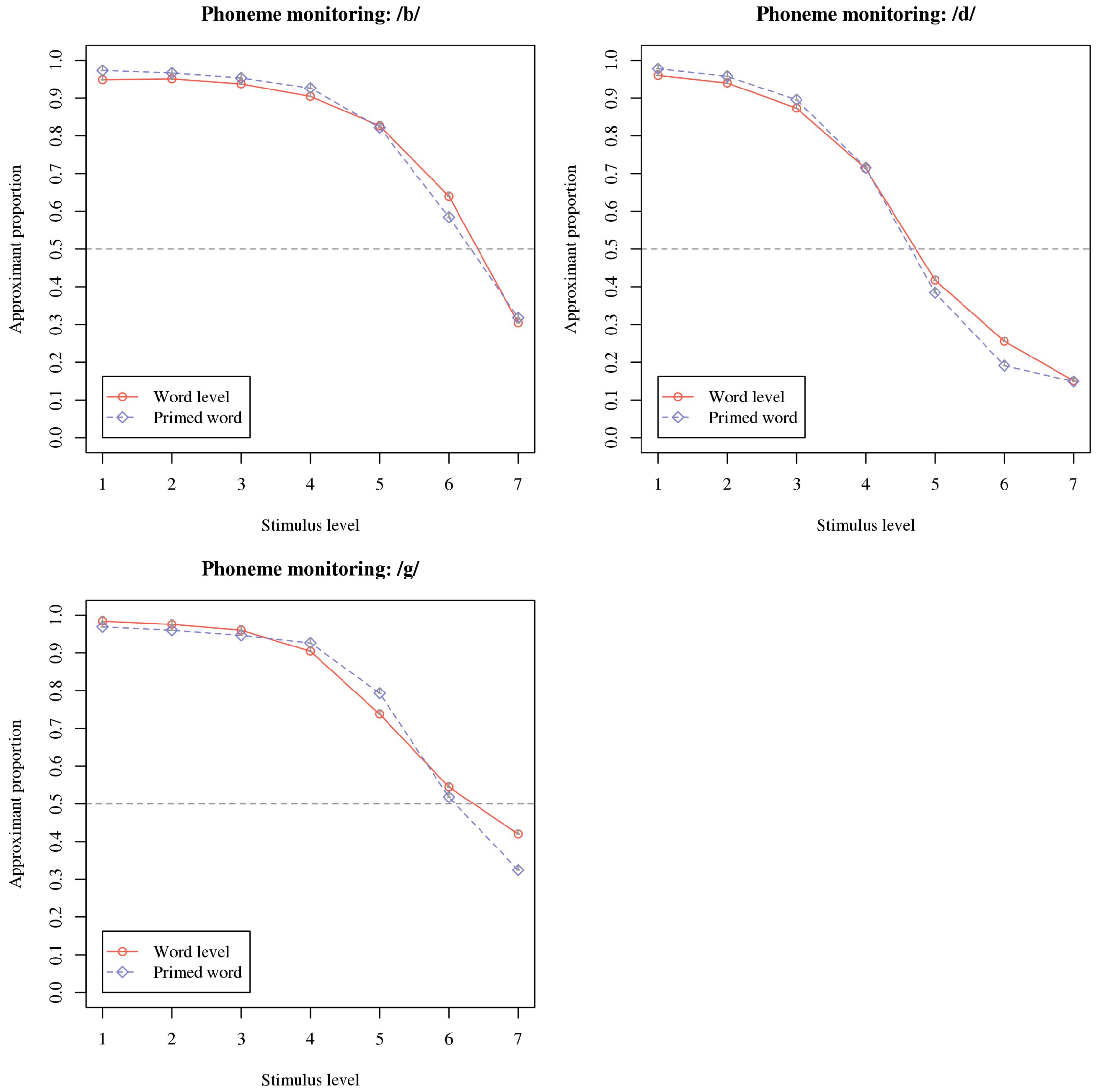

Figure 7.5. Phoneme monitoring results for /b/, /d/ and /g/. Proportion of reported consonant presence is shown as a function of stimulus level, in continua from full approximants to elided variants, in two conditions: word-level and primed word-level (for each consonant, *n* = 6300). 95% confidence interval bars are included.

A GLMM analysis was conducted on the results of phoneme monitoring for /b/. The best fitting model included response as the dependent variable, stimulus level (as a continuous variable) as a main effect, subject as a random factor, and stimulus level and experimental condition as random slopes. The assumption of normality for the residuals of this model was assessed using histograms and quantile-quantile plots; no important deviations from normality were found (skewness: -1.018; excess kurtosis: 2.268). The results of this analysis showed a significant main effect of stimulus level ($\chi^2(1) = 73.689$, $p < 0.001$). The fact that adding experimental condition to the model did not improve its fit can be interpreted as a non-significant main effect. The same can be said of the interaction between stimulus level and experimental condition.

Results for /d/

The results for the word-level condition approximated a cumulative binomial distribution. The first three steps reached values around ceiling, and crossed the 50% chance level between steps 4 and 5, to finally settle close to 20% perception on step 7 (see top-right panel from Figure 7.5). The primed word condition essentially replicated the results from the word-level condition, although slightly higher consonant perception was observed in the first half of the continuum, and slightly lower in the second half. The predictions of a category boundary shift in favour of [ð̞] perception were not met.

A GLMM analysis was conducted on the results of phoneme monitoring for /d/. The best fitting model included response as dependent variable, stimulus level as a main effect, subject as a random factor, and stimulus level and experimental condition as random slopes. The assumption of normality for the residuals from this model was assessed using histograms and quantile-quantile plots; no important deviations from normality were found (skewness: -0.166; excess kurtosis: 1.551). The results showed a significant main effect of stimulus level ($\chi^2(1) = 222.57$, $p < 0.001$), but no main effect of experimental condition or interaction between stimulus level and condition.

Results for /g/

The results for the word-level condition approximated the first half of a cumulative binomial distribution. The first four steps reached values around ceiling and crossed chance level at step 6. Perception fell to around 40% at step number 7 (see bottom panel from Figure 7.5). The primed word condition showed very similar results as the word-level condition. Once more, the predictions of a category boundary shift in favour of [ɣ̞] perception were not met.

A GLMM analysis was conducted on the results of phoneme monitoring for /g/. The best fitting model included response as dependent variable, stimulus level as a main effect, subject as a random factor, and stimulus level and experimental condition as random slopes. No relevant deviations from normality were observed for the residuals from this model (skewness: -0.871; excess kurtosis: 2.715). The results showed a significant main effect of stimulus level ($\chi^2(1) = 228.25$, $p < 0.001$), but no main effect for experimental condition nor a significant interaction between stimulus level and experimental condition.

Summary

Perception of the three target consonants decreased as a function of stimulus level in both experimental conditions, in line with the expectation that less acoustic evidence would be interpreted by listeners, to some extent at least, as absence of the segmental unit (see Figure 7.5). In all cases, consonant perception started close to ceiling and remained at that level until step 3, around which point perception decreased gradually to cross chance level between steps 4 and 5 for /d/, and closer to step 6 for /b/ and /g/. Only for /d/ did results approach floor perception; consonants /b/ and /g/, on the other hand, showed a delay in the perception of elided variants, with the last steps not reaching the lowest levels of perception.

While a main effect of stimulus level was found for the three consonants, no main effect of experimental condition was detected, nor was there an interaction between stimulus level and experimental condition. In summary, there is no statistical evidence for semantic priming effects.

7.4. Discussion

It is useful to begin this discussion by highlighting some differences between the experiments reported here and those from Chapter 6. The most important difference relates to the fact that in these continua the elided end does not constitute a legal alternative Spanish word, but instead it can only be interpreted as the same lexical item with an elided approximant (or as a nonsense word). This difference has important consequences, because in these continua only one lexical item can cause a lexical effect on speech perception (Ganong, 1980). The second important difference is that target lexical items display higher lexical frequencies when compared to the other experiments, and primes were also stronger. Finally, it is relevant to point out that the results are all from word-level continua, and not from segmental conditions, which displayed response patterns that differed considerably from word-level conditions in previous experiments.

The response distributions can be grouped into two distinct patterns: the one described by /b/ and /g/, in which perception never reached floor and more recovery was observed, and the one from /d/, closer to a categorical perception distribution. In the case of /b/ and /g/, the results can be interpreted as evidence of a lexical effect on speech processing, which triggers recovery (Ganong, 1980), although a comparison to a baseline segmental condition would be required to determine this conclusively. This hypothesized lexical effect could be present on the data despite the fact that phoneme monitoring is a primarily auditory task in which listeners can choose to ignore higher levels of lexical processing and provide an auditory response. Although it is not possible to disregard a possible a lexical effect on the results from /d/, it is clear that, if present, it is not as strong as for /b/ and /g/.

One possible explanation as to why /b/ and /g/ displayed stronger (hypothetical) lexical effects and more recovery has to do with the expectations that listeners have with respect to what is normal in production and natural perception (Mitterer & Ernestus, 2006; Janse et al., 2007). For /d/, in which elision is the norm in production and natural perception, a highly lenited or elided variant is a good example of an underlying /d/, and thus the elided end from the continuum could have been interpreted, primarily, as the target word with an elided consonant. In the case of /b/ and /g/, in which elision is less likely in production and natural perception, a highly lenited instance is not necessarily a good representative of the underlying phonological unit (or the episodic cloud), and thus the elided end of the continuum can be interpreted more often as a nonsense word, which would explain a stronger lexical effect and more recovery for these consonants.

No statistically significant differences were found between experimental conditions (word-level and primed word), and thus the statistical evidence does not support the hypothesis of a semantic priming effect, despite some trends that could be interpreted in that direction[21]. These results confirm that the semantic priming technique used in the perception experiments was not able to render a detectable semantic priming effect, even for relatively frequent lexical items and strong primes. It is expected that embedding continua like the ones used here in a sentence level should display clear and strong semantic effects on perception (Ernestus et al., 2002; Kemps et al., 2004). However this method only works when only one lexical interpretation can be extracted from the continuum step, which was not the case in the experiments from Chapter 6.

What lexical access model might account for these results? Any model in which higher levels of speech processing can have an effect on lower levels of perception can account for the lexical effects and recovery results observed in the data. Episodic models in general are not very well suited to explain the results, but Minerva 2 (Hintzman, 1984, 1986; Goldinger, 1998) does accept some type of top-down feedback by assuming that the echo returned after a match contains more information than just a label for the input episode, but rather characteristics from the whole episodic cloud. In the case of abstractionist models, autonomous race models such as RACE (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987), Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997) and Merge (Norris, McQueen, & Cutler, 2000) include parallel lexical processing routes that can account for lexical effects on speech perception, although the exact implementation of this parallel route is not very clear for Shortlist, while in Merge there is feedback from a lexical level to phonemic decision nodes, in what resembles top-down feedback. Other abstractionist models like TRACE (McClelland & Elman, 1986a, 1986b) do include bidirectional connections between different processing levels, and consequently feedback from the lexical levels can affect prelexical levels of perception. Finally, connectionist hybrid models are particularly well suited to explaining the results observed in our experiments. In particular, Goldinger's CLS (Goldinger, 2007) and POLYSP (Hawkins & Smith, 2001; Hawkins, 2003) model lexical feedback to lower levels of perception. The fact that episodes are

21 For the three consonants, but more clearly for /d/, primed conditions showed higher proportion of consonant perception than non-primed conditions before the category boundary, and less perception levels after it. In short, semantic priming seemed to maximize the perceptual contrast between both ends of the continua, without shifting the category boundary.

an integral part of these hybrid models allows them to recruit fine phonetic detail in perception, which in the case of /b d g/ means that episodic clouds should be better represented by the most frequent variants from each consonant, and thus means that they can account for the expectations that listeners have regarding what is expectable in natural production and perception.

Chapter 8
General discussion

8.1. Lenition in Chilean Spanish /b d g/: linking production and perception

Evidence from production showed that lenition and elision are the norm in Chilean Spanish (see Chapter 4). This was true for the three consonants investigated, but particularly so for /b/ and /d/, in which vocalic approximants and elided variants were more common. Approximant consonants displayed a natural continuum of realizations from full approximants to elided variants, which was encoded acoustically as a combination of duration, intensity and F1. These three acoustic variables behaved as expected given previous research: more lenited variants are shorter, less loud and have higher F1 values. Also, the variation occurs in weaker contexts (e.g., intervocalic), in words, in more informal tasks and in more frequent words.

If absence of acoustic evidence is normal for Chilean Spanish approximants of /b d g/, and if there is some advantage in leniting these units, one fair expectation could be that lenition should take place categorically, that is, that units should be either fully present of absent from the signal, but not gradually present. An initial answer to this hypothesis is simply that data suggests otherwise, and that one general property of lenition seems to be that it takes place gradually. A more comprehensive answer has to consider the goal that a strategy like lenition tries to accomplish (e.g., effort reduction) at the same time as the restrictions that the phonetic and phonological systems impose on this strategy (e.g., intelligibility, contrast maintenance) (cf. Lindblom, 1990). Whichever may be the case, approximant consonants of Spanish, and particularly of the Chilean variety, can be said to constitute *low information bearing units*, that is, units from the phonetic system that carry little information about themselves, as a result of their relatively poor acoustic constitution, and due to the fact that they are often degraded or absent from the acoustic signal altogether (another similar example is English schwa). Low information bearing units should consequently be very unreliable to listeners, a hypothesis that was confirmed for /b/ and /d/ in the perception studies (see Chapter 6). This approach is compatible with theories that see lenition as the degradation of the informational complexity of the speech signal (Harris & Urua, 2001), or as a means to decrease the amount by which a segment interrupts the speech flow, in order to make neighbouring prosodic constituents more prominent (Kingston, 2008).

In the perception domain, one key finding was that perception of the segmental condition was different for /b/, /d/ and /g/, and that these differences could be related to how reliable the acoustic information from the approximants was in natural production and perception. These results show that perception in tasks and conditions closer to the auditory domain are affected by what listeners expect and know about production and perception. This has very interesting implications for speech perception in general. If even pre-lexical auditory tasks can be affected by more than only the acoustic cues available to listeners, this means that perception is biased (to some extent) by experience, that is, an acquired bias. Notice that this evidence is not necessarily explained fully by positing prototypes or best exemplars that distort the perceptual space in favour of existing categories, as in the magnet effect (Kuhl, 1991; Iverson & Kuhl, 1995; Kuhl & Iverson, 1995), given that in this case the accumulated experience biases the way in which the presence or absence of a unit is assessed. Moreover, for all experiments from Chapter 6, the absence of the underlying unit also constituted an acceptable and comprehensible linguistic entity, which sets this design apart from previous experiments controlling contextual cues for highly lenited forms (Ernestus, Baayen, & Schreuder, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004; Mitterer & Ernestus, 2006).

Another key finding from the perception domain was that categorical perception increased when lexical effects were present, and that lexical effects were clearer in a task in which post-lexical processing was mandatory (i.e., identification). In conditions in which word-level cues were available, distributions became closer to categorical. In practice, this meant that they became more symmetrical, and that a category boundary was observable for all consonants. While it is a possibility that no lexical effects (as in Ganong, 1980) were present, since two competing lexical items were present and no end of the continua –in Chapter 6 at least– rendered a nonsense word, it is also possible that two lexical effects were taking place at the same time, not biasing the continuum to one interpretation, but instead bringing perception of the approximant consonants closer to equilibrium.

As for semantic priming, weak effects of semantic priming were only detected for /d/, a consonant with particularly unreliable acoustic evidence in natural perception. It seems to be the case that there is a limit to the effects that various cues have in speech perception. Some cues, such as adding a word-level context, have the effect of dramatically changing the way in which listeners perceive a continuum from consonant presence to absence, and the same can be expected of adding sentence level syntactic and semantic cues. Semantic priming, instead, barely had an effect. As discussed earlier, this was probably due to weak effects, not detected in the statistical analyses, or due to a failure to find truly highly frequency words and strong primes. An alternative explanation is that phonetic categorization simply cannot be affected by priming semantic associates, that is, by activating a word that might also activate related items. This explanation is supported by the fact that a semantic priming effect was not found even for highly frequent words and strong semantic primes, as in *herradura* ("horseshoe") as the prime for *caballo* ("horse") (see Chapter 7 for details, in particular section "7.2.2. Stimuli" and Table 7.1). It may be the case that once a word-level context is provided, phonetic categorization is not further affected by activating associated lexical items, and that only one interpretation of the continuum via additional syntactical and semantic context from the sentence level is found.

Results from the discrimination tasks could be interpreted both as evidence for or against categorical perception. Sensitivity to differences between stimuli increased for those stimuli closer to category boundaries, and sensitivity was close to zero for the segmental condition of those consonants with unreliable acoustic evidence (/b/ and /d/). Unless the listeners' accumulated experiences in production and perception are taken into account, there is no reason why the continua from the segmental conditions should not have been discriminated categorically. However, once additional cues set in, discrimination increased considerably for both consonants, although the strongest effect was observed for the consonant with the least reliable acoustic properties (i.e., /d/). This suggests strong feedback from higher levels of speech processing to lower levels of speech perception.

Our results from production and perception suggest a strong link between the two domains. For primarily auditory tasks and conditions at least, perception seems to be conditioned by expectations related to what is normal in the production domain.

Similarly, it is to be expected that speakers produce approximant consonants with their listener in mind, knowing which articulatory efforts will render sufficiently interpretable acoustic results. Different theoretical assumptions result in very different hypotheses regarding the exact nature of this perception-production link, and of the processes involved. For example, if underlying phonological categories are assumed, then part of perception consists in normalizing the input to retrieve phonetic invariants that can then be mapped into underlying phonological units. Production, on the other hand, requires an inverse process in which abstract underlying units are converted into a group of features that need to be encoded articulatory before surfacing as actual articulatory gestures. A link between production and perception such as the one observed in our results would require a set of rules mediating between the acoustic input and its underlying representation, biasing perception in a given direction coherent with experience, or boosting perception of those units lacking sufficient acoustic evidence by relaxing perception thresholds in tasks where postlexical processing is not required. Alternatively, if episodes are posited, linguistic experience becomes organically organized in exemplar clouds. More frequent input becomes better represented and production could be a simple process of retrieving one exemplar and translating it into articulatory gestures. The link between production and perception would thus be encoded in the episodic clouds.

Given that it escapes the explanatory scope of most lexical access models, few of them are explicit about how perception and production might be related. Some hybrid models like Pierrehumbert's ED, describe production as the process of a given phonetic category activating a label, which triggers the selection of a random exemplar from the cloud of exemplars associated with the label to be rendered by articulatory means (Pierrehumbert, 2001, 2002). A radically different approach is developed by the Motor Theory of Speech Perception (MTSP) (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), which proposes that speech perception and production are intimately linked by a common set of processing strategies and representations, involving underlying intended gestures. While production is modelled as a sequence of rules converting implicit motor instructions to neural commands, muscle contractions and then articulatory shapes, perception requires performing analysis-by-synthesis in order to link the acoustic input to underlying articulatory intentions. In order to explain the results obtained in production, intended underlying articulatory gestures would need to encode or be linked to rules determining acceptable degrees of lenition (perhaps, as acceptable target-undershoot thresholds). In the case of perception, higher sensitivity in the analysis-by-synthesis process should be allocated to segments requiring small articulatory gestures.

8.2. Challenges to lexical access models

The results of the production and perception experiments challenge lexical access models at several levels. One clear example comes from the differences that were observed in the perception of /b d g/ in the segmental task in phoneme monitoring and identification. The fact that different consonants showed different trends, with perception failing to reach ceiling for /b/ even when full acoustic evidence was present, and, in the case of /d/, with perception failing to reach floor when evidence was absent, makes it unlikely that mapping the acoustic input to underlying phonological units is driving the perceptual process, even if underspecified features are possible, as in Cohort (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987, 1989; Lahiri & Marslen-Wilson, 1991, 1992), or if gradual featural matching is possible, as in the Fuzzy Logical Model of Perception (Oden & Massaro, 1978; Massaro & Oden, 1980). This is so because in abstractionist models of lexical access underlying units are actually indifferent to the type of variation resulting from processes such as lenition, because that information is not encoded into them and is lost at early normalization stages (Ernestus, 2014). Moreover, normalization ought to be particularly strong in tasks in which no other information is able to disambiguate the input, such as in phoneme monitoring of the segmental condition.

These observations seem to constitute evidence in favour of an episodic memory account for the data from the perception experiments conducted here, since they are able to store the fine phonetic detail originating from perceptual and production experience. The influence of episodes ought to be clearer in an auditory task, in which the acoustic information itself is scrutinized. Conversely, in tasks in which top-down feedback becomes available, the relative strength of episodes should decrease, as in the identification task. When additional cues become available, and the listener can tap into top-down information, postlexical feedback seems to override the effects of episodic traces seen in the segmental task, where phonetic experience had a clearer influence in perceptual results. Interactive episodic models such as Minerva 2, and interactive hybrid models such as Goldinger's CLS and POLYSP are able to account for these results.

As discussed above, the fact that expectations regarding what is normal in natural production and perception have an effect on prelexical stages of lexical access suggests that episodes play a role in perception. The evidence is less clear regarding whether underlying abstract representations are also required or not. When additional cues become available, in particular word-level cues, lexical effects set in, overriding the effects that episodes had in tasks and conditions where post-lexical processing was not mandatory. Both episodes and underlying abstract phonological representations can account for these effects. In the case of episodes, the amount of overlap between the label of an episodic cloud and the input, considerably larger than in the segmental condition, ought to facilitate lexical effects and recovery. A very similar explanation can be elaborated for a chain of underlying phonological units: as soon as lexical level information plays a role, underlying units should be able to tolerate a degree of mismatch for a segment given clear evidence for the rest.

Comparing perception results to models' predictions

A summary table is provided (see Table 8.1) to relate the outcomes of the speech perception experiments from Chapter 6 and Chapter 7 and the predictions made by broad families of models of speech perception and lexical access.

Table 8.1. Summary table of the results of the perception experiments and their relation to the predictions from abstractionist, episodic and hybrid models of speech perception and lexical access.

<table>
<tr><th></th><th colspan="3">Predictions of models of speech perception and lexical access</th></tr>
<tr><th>Observed result</th><th>Abstractionist</th><th>Episodic</th><th>Hybrid</th></tr>
<tr><td>Chapter 6: The effect of adding contextual and semantic cues was practically null in the perception of units well represented by acoustic evidence in natural perception (see evidence for /g/).</td><td rowspan="2">Models without partial feature matching (e.g., RACE, Merge) predict categorical perception for consonants well represented acoustically. Matches results: Yes.

Models with partial feature matching (e.g., FLMP) predict categorical perception. Matches results: Yes.

In RACE models (i.e., Race, Shortlist, Merge), the prelexical route should form the basis for lexical access for reliable acoustic input in tasks focusing in prelexical speech processing. Similarly, in connectionist models (i.e., TRACE), postlexical feedback ought to be relatively unimportant. Matches results: Yes.</td><td>The perception of units well represented acoustically in exemplars should not be affected (to a great extent) by additional contextual and semantic cues. Matches results: Yes.</td><td>The perception of units well represented acoustically in exemplars should not be affected by additional contextual and semantic cues, although top-down feedback cannot be ruled out in some models (e.g., CLS and POLYSP). Matches results: Yes.</td></tr>
<tr><td>Chapter 6: Adding contextual and semantic cues increased perception and phonological recovery in the perception of units relatively well represented by acoustic evidence in natural perception (see evidence for /b/).</td><td>The perception of units relatively well represented acoustically in exemplars should not be affected (to a great extent) by additional contextual and semantic cues. Matches results: No.</td><td>The perception of units relatively well represented acoustically in exemplars could be affected by additional contextual and semantic cues, increasing perception; top-down feedback cannot be ruled out in some models (e.g., CLS and POLYSP). Matches results: Yes.</td></tr>
<tr><td>Chapter 6: Listeners recovered phonological units for which the acoustic evidence was scarce or null in experimental settings in the perception of units poorly represented by acoustic evidence in natural perception. Adding contextual and semantic cues removed this effect (see evidence for /d/).</td><td>Models without partial feature matching (e.g., RACE, Merge) predict categorical perception for consonants poorly represented acoustically. Matches results: No.

Models with partial feature matching (e.g., FLMP) predict high variability perception of units poorly represented acoustically as well as high perception of partially matching input. Matches results: Yes.

In models with a lexical route (e.g., RACE, Shortlist, Merge), lexical access can be</td><td>When the majority of episodes consist of scarce acoustic representations, the perception of a segmental condition should be relatively high for all continuum steps from approximant to elision (scarce evidence is sufficient to trigger perception). Matches results: Yes.

Additional contextual and semantic cues can remove this effect if lexical-sized matches are possible for word-level stimuli. Matches results: Yes.</td><td>In hybrid models in which post-lexical feedback to lower levels of speech perception is possible (e.g., CLS and POLYSP), the perception of units represented in episodes by scarce acoustic information ought to be relatively high in the segmental conditions, and additional contextual and semantic cues ought to bring perception closer to categorical perception. Matches results: Yes.

In models in which post-lexical feedback is not possible (e.g., Pierrehumbert's ED),</td></tr>
</table>

Observed result	Predictions of models of speech perception and lexical access: Abstractionist	Episodic	Hybrid
	achieved by the lexical route. Similarly, in connectionist models (i.e., TRACE), lexical feedback ought to be relatively strong. *Matches results*: **Yes**.		matching acoustically poor episodes to underlying phonological sized categories ought to be difficult in the segmental condition and ought to increase when additional cues are available. *Matches results*: **No**.
Chapter 6: Very weak effects from semantic priming were only detected for a consonant with particularly unreliable acoustic evidence in natural perception (see evidence for /d/).	Models which explicitly deal with semantic priming (i.e., Cohort) predict that priming should facilitate the activation of the primed candidate, particularly so in the event of ambiguous input. *Matches results*: **Yes**.	Semantic priming should not affect categorical perception. *Matches results*: **Yes**.	Semantic priming should not affect categorical perception. *Matches results*: **Yes**.
Chapter 6: Evidence for categorical perception increased when lexical effects from two comparable lexical competitors were present.	Pure bottom-up models (e.g., Shortlist) predict no increase in the perception of a unit when two comparable lexical competitors become available. *Matches results*: **No**. Race (e.g., RACE, Cohort, Merge) and connectionist models (e.g., TRACE) predict that categorical perception can be affected by post-lexical levels of processing, particularly so for ambiguous input. Having two comparable lexical competitors ought to facilitate categorical perception. *Matches results*: **Yes**.	Continua well represented acoustically in episodes (see results from /g/) should be perceived categorically in word-level conditions. *Matches results*: **Yes**. Continua for units less well represented acoustically in episodes should display a category boundary shift in favour of the approximant interpretation of continua in word-level conditions. *Matches results*: **No**.	In connectionist hybrid models (e.g., CLS and POLYSP), perception can become closer to categorical when the lexical effects from two comparable lexical competitors are present. This ought to be clearer for consonants represented poorly in the acoustic signal. *Matches results*: **Yes**. In non-connectionist models (e.g., Pierrehumbert's ED), categorical perception should increase when word-sized stimuli become available, due to better matches to episodes. *Matches results*: **Yes**.
Chapter 6: Lexical effects were clearer in a task where post-lexical processing was mandatory (identification), and less clear in a task in which listeners could choose to ignore post-lexical processing (phoneme monitoring).	Pure bottom-up models (e.g., Shortlist) predict no differences between tasks with different locus of processing (pre-lexical versus post-lexical). *Matches results*: **No**. Race (e.g., RACE, Cohort, Merge) and connectionist (e.g., TRACE) models predict	In strong episodic models (e.g., LAFS), lexical effects should be similar in tasks requiring mandatory post-lexical processing and those in which listeners can ignore post-lexical processing. *Matches results*: **No**.	In non-connectionist models (e.g., Pierrehumbert's ED), lexical effects ought to be similar in tasks requiring mandatory post-lexical processing and those in which listeners can ignore higher levels of lexical processing.

Observed result	Predictions of models of speech perception and lexical access: Abstractionist	Episodic	Hybrid
	that lexical effects ought to be stronger in tasks requiring post-lexical processing. *Matches results*: **Yes**.	In episodic models in which some type of abstraction is possible (e.g., Minerva 2), lexical effects could be stronger in tasks requiring mandatory post-lexical processing. *Matches results*: **Yes**.	In connectionist models (e.g., CLS and POLYSP), lexical effects ought to be clearer in tasks and conditions requiring mandatory post-lexical processing. *Matches results*: **Yes**.
Chapter 6: Sensitivity to stimulus differences increased as the amount of acoustic evidence decreased in a continuum.	Discrimination maxima ought to coincide with category boundary crossings from identification tasks (although it is unclear how this applies to the comparison of the presence of a unit with its absence). *Matches results*: **Partially**.	Assuming that episodes aggregate naturally depending on their similarity, and that frequent and recent episodes are better represented in episodic clouds, sensitivity to stimulus differences ought to coincide with category boundary crossings from identification. *Matches results*: **Partially**.	Under the same assumptions than episodic models, sensitivity to stimulus differences ought to coincide with category boundary crossings. *Matches results*: **Partially**.
Chapter 6: Sensitivity to stimulus differences was generally low for consonants with unreliable acoustic cues in natural perception.	Evidence for categorical discrimination ought to be clearer for consonants well represented acoustically in the signal, and less clear for consonants poorly represented acoustically. *Matches results*: **Yes**.	Discrimination sensitivity ought to be lower for acoustic input poorly represented in episodes (i.e., that from /d/ and perhaps /b/) as opposed to input well represented in episodes (i.e., /g/), particularly in conditions without word-level cues. *Matches results*: **Yes**.	Discrimination sensitivity ought to be lower for acoustic input poorly represented acoustically in episodes, and higher for input well represented, since mapping episodes to underlying phonological categories requires extracting its features first. *Matches results*: **Yes**.
Chapter 6: Sensitivity to stimulus differences increased when semantic cues were provided.	Pure bottom-up models (i.e., Shortlist) would predict no sensitivity improvement for conditions in which word-level cues are available. *Matches results*: **No**. In race (i.e., RACE, Cohort and Merge) and connectionist models (i.e., TRACE) post-lexical processing can provide the basis for perception, and thus it might improve discrimination. *Matches results*: **Yes**.	Better matches between acoustic input and episodes ought to occur in word-level stimuli. Discrimination sensitivity should increase once semantic cues are provided. *Matches results*: **Yes**.	Better matches between acoustic input and episodes ought to occur in word-level stimuli for all models. Discrimination sensitivity should increase when semantic cues are provided. *Matches results*: **Yes**.
Chapter 7: Stronger lexical effects and phonological recovery in favour of words was found in the perception of units with	Pure bottom-up models (e.g., Shortlist) predict no lexical effects in ambiguous input, such as that from the elided	In continua from words to nonsense words for consonants well represented acoustically in episodes (i.e., /b/ and /g/),	In connectionist hybrid models, lexical effects ought to be present for continuum steps in which the acoustic input is

Observed result	Predictions of models of speech perception and lexical access Abstractionist	Episodic	Hybrid
more reliable acoustic evidence in natural perception in continua from words to nonsense words (see evidence from /b/ and /g/).	endpoints of continua from /b/ and /g/. *Matches results*: **No**. Race and connectionist models (e.g., RACE, Cohort, Merge; TRACE) predict lexical effects in categorical perception of ambiguous input, such as that from the elided endpoints of continua. *Matches results*: **Yes**.	only words are well represented by episodes, and thus lexical effects should be observed. *Matches results*: **Yes**.	ambiguous, particularly when extracting underlying phonological-sized units is difficult. *Matches results*: **Yes**.
Chapter 7: No evidence of lexical effects and weaker evidence of phonological recovery was found in the perception of units with less reliable acoustic evidence in natural perception (see evidence from /d/). In these stimuli, elided endpoints also constitute legal instances of the word target.	Pure bottom-up models (e.g., Shortlist) predict no lexical effects in ambiguous input. *Matches results*: **Yes**. Race and connectionist models (e.g., RACE, Cohort, Merge; TRACE) predict strong lexical effects in categorical perception of ambiguous input. *Matches results*: **Yes**.	In continua from words to non-words for consonants poorly represented acoustically in episodes (i.e., /d/), the absence of acoustic evidence is a good representative of the lexical-sized episode, and thus perception ought to approach categorical perception. *Matches results*: **Yes**.	As in episodic models, the absence of acoustic evidence is a good representative of the lexical-sized episode, and thus perception should approach categorical perception. In this case, matching to underlying phonological units is particularly difficult. *Matches results*: **Yes**.
Chapter 7: No differences were observed between conditions word-level and semantic priming in continua from words to nonsense words (which used high-frequency words and strong primes).	Models which explicitly deal with semantic priming (i.e., Cohort) predict that priming should facilitate the activation of the primed candidate, particularly so in the event of ambiguous input. *Matches results*: **No**.	Semantic priming should not affect categorical perception. *Matches results*: **Yes**.	Semantic priming should not affect categorical perception. *Matches results*: **Yes**.

8.3. Limitations and future research

8.3.1. Temporal aspects of the word-recognition process

A modified version of the phoneme monitoring task was used in Chapter 6 and Chapter 7 to explore the differences in perception between phoneme monitoring, a primarily auditory task, and identification, in which lexical labels are processed lexically before listeners can provide a response. In the original phoneme monitoring paradigm, reaction times are also collected, and are interpreted along with proportion of response (or response correct, if that is the case) in order to address hypotheses involving cognitive effort, recovery, and temporal aspects of the word-recognition process (e.g., Foss, 1969; Cutler & Darwin, 1981; Dijkstra, Roelofs, & Fieuws, 1995). This task requires that listeners provide their responses as fast as possible, an instruction that was not given to participants in these experiments. Since the modifications enabling the collection of reaction times are relatively simple, the failure to implement them could be seen as a limitation of the present dissertation.

Collecting data about the time course of lexical access would allow to obtain relevant information regarding hypotheses about the perception of approximant variants of /b d g/ to be collected. To give just one example, it is well established that exposure to ambiguous input produces longer reaction times (Foss, 1970; Foss & Jenkins, 1973; Swinney & Hakes, 1976). Obtaining reaction times for continua from approximants to elided variants would provide an additional dimension to complement the observations made about categorical perception, confirming that stimuli near the category boundary are considered ambiguous by participants, despite the fact that some of those variants are the norm in production (see Chapter 4). This type of evidence would also allow evaluation of the predictions of lexical access models. While lexical access models like RACE make very specific predictions for this scenario, stating that the lexical route should provide a response first most of the time (Cutler & Norris, 1979; Cutler, Mehler, Norris, & Segui, 1987), it is less clear how many other models fare on this regard.

Besides phoneme monitoring, in which reaction time is one of two sources of information (the other being the detection of the segment being monitored), there are other psycholinguistic tasks specifically designed to obtain information about how perception unfolds over time, such as the *gating task* and the *speech shadowing task*. In the gating task paradigm (see Grosjean, 1980; Cotton & Grosjean, 1984), sections of word-level stimuli are presented to participants from the beginning of the word up to a certain point in successive “passes”, in which the stretch of the word is increased in regular intervals until the word has been completely revealed. For each iteration, participants are asked to identify and report (often in writing) the word that they are being presented with. Given its design, this task is able to determine when a word can begin to be guessed relative to its competitors, when it becomes a unique candidate and when it is actually recognized. This task can thus be used to determine whether a word can be retrieved before the evidence for it has been fully presented (Marslen-Wilson, 1987), or to explore what happens in perception at the point at which a minimal pair becomes disambiguated (e.g., Sebastián-Gallés & Soto-Faraco, 1999). More generally, this task allows exploration of hypotheses about on-line processing of spoken language, which has a direct bearing on theories of phonological recovery and lexical effects on perception, and of course on the predictions that lexical access models make regarding these issues. Lastly, the gating task is directly relevant to any model in which the left-to-right nature of speech perception is a central element of the architecture, as it is the case for Cohort and Shortlist.

In the case of the speech shadowing task (see Chistovich, 1960; Marslen-Wilson, 1973), listeners are required to repeat continuous speech as they hear it, and the time difference between the input and the time of its repetition –the response latency– is registered at regular intervals. This task has been used to explore the effects of several variables on speech perception; for example, word frequency (Marslen-Wilson, 1985; Radeau & Morais, 1990) and word status (Marslen-Wilson, 1985). Although the gating task seems more appropriate to study similar stimuli to the one used in perception experiment here, speech shadowing could also enable exploration of the properties of on-line speech processing for approximants of Spanish.

8.3.2. Acquisition and L2 learning

Given that Chilean Spanish spirant approximant consonants are poorly represented acoustically in production (see Chapter 4), it might be the case that their acquisition is relatively difficult when compared to that of other allophones from the same series that are better represented by acoustic evidence (e.g., [b d g]), and relative to the members of other series that display considerably less variation (e.g., /p t k/). Unfortunately, very little is known about the acquisition of approximant consonants in Spanish, since most acquisition studies tend to refer to the acquisition of phonological units as a whole (e.g., “the acquisition of /d/”) instead of to particular classes of allophones. For Mexican Spanish, the only variant for which research about the acquisition of approximants could be found, it has been shown that acquisition of [β̞ ð̞ ɣ̞] occurs relatively early in development –around 1 year and 7 months (Macken, 1979)–, and that voiced stops are acquired and produced before approximants, although this was only investigated in word-initial contexts (Macken & Barton, 1980). In the case of Chilean Spanish, a variety with considerably higher degrees of lenition, no information exists about the acquisition of approximant allophones[22].

The question regarding the acquisition sequence of the allophonic variants of the /b d g/ series has important implications for some of the main topics developed in this thesis. For example, it might be the case that infants produce voiced stops first not because they are better representatives of underlying phonological units, but on account of their auditory saliency (i.e., prominence), or, similarly, on account of the unreliability of the acoustic properties of approximants. Investigating infant-directed speech might also provide interesting insights, regardless of the specific sequence of acquisition, since it has been shown that some contrasts of the phonetic system are maximized when speech is addressed to infants (Kuhl et al., 1997), which should have an effect on the variants that infants perceive during early stages of acquisition.

While little is known about the acquisition of [β̞ ð̞ ɣ̞], it is clear that adult native speakers have no problems perceiving these segments in conversational speech, where an array of contextual information is available to listeners to aid in coping with highly lenited or elided units. This contrasts sharply with the difficulties that L2 learners have been shown to have in producing Spanish [β̞ ð̞ ɣ̞]. For example, in the case of English L1 speakers, subjects fail to produce native-like realizations of Spanish approximants in most contexts (Zampini, 1994; Elliott, 1997). This is so, at least partly, because spirant approximants are not allophones of English /b d g/, and because some allophones of

22 By the age of 3 years and 5 months, 70% of children have acquired /b d g/ according to a study by Vivar and León Valdés (2009). No distinction is made between allophonic realizations of /b d g/.

Spanish /b d g/ map into different phonemes of English, as is the case for Spanish [β̞] and [v] of /b/, and [d] and [ð̞] of /d/ (Face & Menke, 2009).

Given these differences, it would be interesting to investigate whether English native speakers learning Spanish as an L2 are particularly challenged by Chilean Spanish approximants, given their natural variation in which highly lenited and elided variants are the norm, at least for /b/ and /d/. A preliminary hypothesis based on my own observations is that, indeed, L2 speakers face significant challenges in producing and recovering underlying units from approximant variants of /b d g/. In order to investigate this, some of the methodologies from this study could be adapted to explore how L2 speakers with different levels of proficiency produce and perceive spirant approximants of Chilean Spanish. In the production domain, it is likely that L2 speakers will generalize a deletion rule eliding most instances where open and vocalic approximants would be expected for L1 speakers. In the perception domain, category boundaries in continua from approximants to elided variants would probably be less well defined than for L1 listeners, and L2 listeners would also probably show lower levels of perception of approximants than L1 listeners (i.e., they would perceive most instances as absent from the signal). L2 listeners likely have less robust category representations of their L2 phonemes, and therefore are less well equipped to deal with highly lenited forms where there is little acoustic information.

8.3.3. The articulatory domain

As mentioned briefly in Chapter 2, very little is known about the articulatory characteristics of the approximants of Spanish /b d g/. Some recent experiments have begun to explore this field. First, there is electropalatographical evidence from /d/ showing that high constriction levels are found for this consonant after laterals and nasals (Hualde, Shosted, & Scarpace, 2011), and that approximant realizations surface more often after /a/ and /r/ (Hualde, Simonet, Shosted, & Nadeu, 2010). Second, studies of electromagnetic articulometry have enabled reliable intensity constriction degree correlates for /b/ to be determined (Parrel, 2010), and have also shown that [b d g] have complete closures during articulation (Parrel, 2011). Finally, evidence from real-time magnetic resonance imaging has shown that spirantization in /d/ involves less movement of the tongue body than less reduced variants (Parrel, 2012).

Besides secondary sources of information such as audio-visual cues for place of articulation for variants of /b/ (e.g., Sadowsky, 2010), no actual articulatory study has been conducted for Chilean Spanish thus far. Studying the articulatory domain for this variety, one in which particularly high degrees of lenition exist, is of particular interest, because a complete correlation between the articulatory and acoustic domains is not always warranted (Lawson, Scobbie, & Stuart-Smith, 2011). While a strong correlation between elision and the absence of accompanying articulatory gestures is certainly the most likely scenario for Chilean Spanish, without exploring the articulatory domain it is not possible to establish this with certainty. For example, it may be the case that realizations classified as fully elided variants because no acoustic evidence was found for them may still display vestigial articulatory gestures. Finding evidence along these lines would have interesting implications for a number of related domains. For example, it may be the case that producing an articulatory gesture without an audible correlate aids L1 infants at acquiring and even perceiving highly lenited and elided approximants. Evidence in this direction would lend support to theories of speech perception and production that place articulatory gestures at the centre of their theoretical assumptions, as it is the case for the MTSP and Articulatory Phonology (Browman & Goldstein, 1992a, 1992b).

8.4. Concluding remarks

This dissertation investigated lenition in the production and perception of Chilean Spanish approximants of /b d g/ in order to explore hypotheses related to phonological recovery, lexical effects on speech perception and lexical access. Results from the production domain showed that Chilean Spanish displays high levels of lenition and elision, that continua from full approximants to elided variants exist naturally in production, and that differences exist regarding the extent to which lenition and deletion affect /b/, /d/ and /g/. In perception, results showed that increasing the amount of cues from a minimal phonetic context to word-level acoustic and semantic cues had an effect on speech perception, bringing responses closer to distributions in agreement with a categorical perception account. Taken together, the results of production and perception were interpreted as indicative of a link between the two domains, given that listeners' experience regarding what is expectable in natural production and perception seemed to have an effect on lower levels of speech processing. These results are better accounted for by lexical access models positing both episodic memory and feedback from lexical to lower levels of speech processing.

References

Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Haskins Laboratories Status Report on Speech Research*, *29*(30), 15-25.

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, *116*, 3099–107.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike* (pp. 199-213). Springer: New York.

Almeida, M., & Pérez Vidal, C. (1991). Datos acústicos de las consonantes fricativas canarias. *Revista de Filología de la Universidad de La Laguna*, *10*, 7-14.

Amastae, J. (1989). The intersection of s-aspiration/deletion and spirantization in Honduran Spanish. *Language Variation and Change*, *1*(02), 169-183.

Amo, L., Visser, M. E., & Oers, K. V. (2011). Smelling out predators is innate in birds. *Ardea*, *99*(2), 177-184.

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baković, E. J. (1994). Strong onsets and Spanish fortition. *MIT Working Papers in Linguistics*, *23*, 21–39.

Barik, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech*, *20*(2), 116-126.

Barlow, J. A. (2003). The stop-spirant alternation in Spanish: Converging evidence for a fortition account. *Southwest Journal of Linguistics*, *22*(1), 51-86.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4

Bauer, L. (2008). Lenition revisited. *Journal of Linguistics*, *44*(03), 605–624.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*(2), 1001-1024.

Bello, A. (1940 [1834]). Advertencias sobre el uso de la Lengua Castellana. Dirigida a los padres de familia, profesores de los colegios y maestros de escuela. In A. Alonso & R. Lida (Eds.), *El español en Chile. Trabajos de Rodolfo Lenz, Andrés Bello y Rodolfo Oroz* (pp. 49-76). Buenos Aires: Facultad de Filosofía y Letras de la Universidad de Buenos Aires.

Berglund, Å. M., & Nyholm, N. E. I. (2011). Slow improvements of metal exposure, health-and breeding conditions of pied flycatchers (Ficedula hypoleuca) after decreased industrial heavy metal emissions. *Science of the Total Environment*, *409*(20), 4326-4334.

Bickford, A. C., & Floyd, R. (2006). *Articulatory Phonetics: Tools for Analyzing the World's Languages*. SIL International.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.17, retrieved 20 August 2015 from http://www.praat.org/

Boley, J., & Lester, M. (2009, October). Statistical analysis of abx results using signal detection theory. In *Audio Engineering Society Convention 127*. Audio Engineering Society.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135.

Borland Delorme, K. E. (2004). La variación y distribución alofónica en el habla culta de Santiago de Chile. *Onomázein*, *2*(10), 103-115.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.

Browman, C. P., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

Browman, C. P., & Goldstein, L. (1992b). “Targetless” schwa: an articulatory analysis. *Papers in laboratory phonology II: Gesture, segment, prosody*, 26-56.

Brown, E. (2011). Paradigmatic peer pressure: Word-medial, syllable initial /s/ lenition in Dominican Spanish. In *Selected proceedings of the 5th conference on laboratory approaches to Romance phonology* (pp. 46-58).

Brown, E. (2013). Word Classes in Studies of Phonological Variation: Conditioning Factors or Epiphenomena?. In *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 179-186).

Burton, M. W., & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1230.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (eds.), *Usage-based models of language* (pp. 65-85). Stanford: CSLI.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261-290.

Bybee, J. (2003). *Phonology and language use*. Cambridge University Press.

Carrasco, P., Hualde, J. I., & Simonet, M. (2012). Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica*, *69*(3), 149-179.

Cepeda, G. (1991). *Las consonantes de Valdivia*. Valdivia: Universidad Austral de Chile.

Cepeda, G. (1994). Las consonantes del español de Valdivia (Chile). Los procesos de reforzamiento y debilitamiento fonológicos. *Estudios Filológicos*, *29*, 39-61.

Cepeda, G. (2001). Estudio descriptivo del español de Valdivia, Chile. *Estudios Filológicos*, *36*, 81-97.

Cepeda, G., & Poblete, M. T. (1993). Retención y elisión de /β/ y /ð/ en sufijos y morfemas radicales. *Estudios Filológicos*, *28*, 87-96.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456-465.

Chistovich, L. A. (1960). Classification of rapidly repeated speech sounds. *Akusticheskii Zhurnal*, *6*, 392-398.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*(4), 466-485.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics*, *27*(2), 207-229.

Cid Uribe, M., & Céspedes Morales, M. (2008). Rasgos de simplificación en el habla rural de dos localidades de Chile: descripción fonotáctica y discursiva. *Literatura y Lingüística*, *19*, 197-210.

Colantoni, L., & Marinescu, I. (2010). The scope of stop weakening in Argentine Spanish. In *Selected proceedings of the 4th conference on laboratory approaches to Spanish phonology* (pp. 100-114). Somerville, MA: Cascadilla.

Cole, J., Hualde, J. I., & Iskarous, K. (1999). Effects of prosodic and segmental context on /g/-lenition in Spanish. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of the Fourth International Linguistics and Phonetics Conference* (pp. 575–589). Thessaloniki.

Cole, R. A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same-or different-voiced letters. *The Quarterly Journal of Experimental Psychology*, *26*(1), 1-7.

Connine, C. M. (1990). Effects of sentence context and lexical knowledge during speech processing. In G. Altmann (Ed.), *Computational and psycholinguistic approaches to language processing*, pp. 281–294. Cambridge, MA: MIT Press.

Connine, C. M., & Clifton Jr, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 291.

Contreras, C. (1993). El castellano rural de Osorno, Chile (A través de textos orales). *Estudios Filológicos*, *28*, 123-135.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics*, *35*(1), 41-48.

Creelman, C. D., & Macmillan, N. A. (1979). Auditory phase and frequency discrimination: a comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 146.

Croissant, Y. (2013). *mlogit: multinomial logit model*. R package version 0.2-4. http://CRAN.R-project.org/package=mlogit

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science*, *12*(5), 725-731.

Cutler, A. (1989). Auditory lexical access: where do we start?. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 342–356). Cambridge, MA: MIT Press.

Cutler, A. (1998). The recognition of spoken words with variable representations. In *Proceedings of the ESCA workshop on sound patterns of spontaneous speech* (pp. 83–92).

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics*, *29*(3), 217-224.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory phonology*, *10*, 91-111.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*(2), 141-177.

Danesi, M. (1982). The description of Spanish /b, d, g/ revisited. *Hispania*, *65*(2), 252-258.

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. *Word*, *8*(3), 195-210.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology*, *49*(2), 264.

Duffy, M. A., Cáceres, C. E., Hall, S. R., Tessier, A. J., & Ives, A. R. (2010). Temporal, spatial, and between-host comparisons of patterns of parasitism in lake zooplankton. *Ecology*, *91*(11), 3322-3331.

Dunn, O. J. (1959). Estimation of the medians for dependent variables. *The Annals of Mathematical Statistics*, 192-197.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*(293), 52-64.

Eddington, D. (2011). What are the contextual phonetic variants of in colloquial Spanish?. *Probus*, *23*(1), 1-19.

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, 95-108.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(02), 143-188.

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, *40*(4), 481-92.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27-41.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1), 162-173.

Escure, G. (1977). Hierarchies and phonological weakening. *Lingua*, *43*, 55–64.

Face, T. L., & Menke, M. R. (2009). Acquisition of the Spanish voiced spirants by second language learners. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 39-52). Somerville, MA: Cascadilla Proceedings Project.

Foley, J. (1970). Phonological distinctive features. *Folia Linguistica*, *4*(1-2), 87-92.

Forster, K. I. (1989). Basic issues in lexical processing. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 75–107). Cambridge, MA: MIT Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627-635.

Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, *29*(2), 137-158.

Foss, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, *8*(4), 457-462.

Foss, D. J. (1970). Some effects of ambiguity upon sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *9*(6), 699-706.

Foss, D. J., & Jenkins, C. M. (1973). Some effects of context on the comprehension of ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 577-589.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. J*ournal of the Acoustical Society of America*, *101*(6), 3728-3740.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of “new” and “old” words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489-504.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression*, Second Edition. Thousand Oaks, CA: Sage.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception and performance*, *10*(4), 526.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268-294.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America*, *124*(2), 1234-1251.

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition*, *17*(2), 134-140.

Fucikova, E., Drent, P. J., Smits, N., & Van Oers, K. (2009). Handling stress as a measurement of personality in great tit nestlings (Parus major). *Ethology*, *115*(4), 366-374

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789-806.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*(3), 447-463.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 49-54).

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*(1), 133-159.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, *103*(5), 2677-2690.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, *99*(2), 221.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Chicago Linguistic Society* (Vol. 35, pp. 151-166).

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267-283.

Hammond, R. M. (1976). Phonemic restructuring of voiced obstruents in Miami-Cuban Spanish. In F. M. Aid, M. C. Resnick & B. Saciuk (Eds.), *1975 Colloquium on Hispanic linguistics*. Paper presented at the 1975 Colloquium on Hispanic Linguistics (pp. 42-51). Washington D.C.: Georgetown University Press.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 1-52). New York, NY: Cambridge University Press.

Harper, D. (2014). *An Analysis of Perceptual Factors in the Evolution of Spanish Approximants* (Unpublished doctoral thesis). University of Colorado Boulder, Colorado, USA.

Harris, J. W. (1969). *Spanish phonology*. Massachusetts Institute of Technology.

Harris, J., & Urua, E.-A. (2001). Lenition degrades information: consonant allophony in Ibibio. *Speech, Hearing and Language: Work in Progress, 13*, 72–105.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*(3), 373-405.

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics*, *13*, 99-188.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099-3111.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Hodge, F. S., Colton, R. H., & Kelley, R. T. (2001). Vocal Intensity Characteristics in Normal and Elderly Speakers. *Journal of Voice*, *15*(4), 503-511.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*, 3059-3071.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge University Press.

Hualde, J. I., Shosted, R., & Scarpace, D. (2011). Acoustics and articulation of Spanish /d/ spirantization. In *Proceedings of the International Congress of Phonetic Sciences XVII* (pp. 906-909).

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2), 301-329.

Hualde, J. I., Simonet, M., Shosted, R., & Nadeu, M. (2010, March). Quantifying Iberian spirantization: Acoustics and articulation. In *40th Linguistic Symposium on Romance Languages, Seattle, WA* (pp. 26-28).

Ingram, J. C. (1989). Connected speech processes in Australian English. *Australian Journal of Linguistics*, *9*(1), 21-49.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553-562.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, *128*(2), 839-850.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language variation and change*, *21*(02), 233-256.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, *22*(2), 161-200.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29-54). Tokyo, Japan: The National International Institute for Japanese Language.

Jongman, A., Wang, Y., & Kim, B. H. (2003). Contributions of semantic and facial information to perception of nonsibilant fricatives. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1367-1377.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252-1263.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, *45*, 229-254.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2004). Domain-initial articulatory strengthening in four languages. *Papers in laboratory phonology VI: Phonetic interpretation*, 145-163.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*(1), 117-127.

Kendall, T., & Thomas, E. R. (2014). *vowels: Vowel Manipulation, Normalization, and Plotting*. R package version 1.2-1. http://CRAN.R-project.org/package=vowels

Kingston, J. (2008). Lenition. In L. Colantoni & J. Steele (Eds.), *Selected proceedings of the 3rd conference on laboratory approaches to Spanish phonology*. Paper presented at the 3rd conference on laboratory approaches to Spanish phonology, Victoria College, University of Toronto, Toronto, 8-10 September (2006). Somerville, MA: Cascadilla Proceedings Project.

Kirchner, Robert. (1998). *An effort-based account of consonant lenition* (Unpublished doctoral thesis). University of California at Los Angeles, Los Angeles, USA.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279-312.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*(2), 820-857.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In *Speech production and speech modelling* (pp. 69-92). Springer Netherlands.

Kotrlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and Interpreting Effect Size in Quantitative Agricultural Education Research. *Journal of Agricultural Education*, *52*(1), 132-142.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93-107.

Kuhl, P. K., & Iverson, P. (1995). Chapter 4: Linguistic Experience and the “Perceptual Magnet Effect,”. *Speech perception and linguistic experience: Issues in cross-language research*, 121-154.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684-686.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-29. http://CRAN.R-project.org/package=lmerTest

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*(01), 97-120.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonology, Phonetics, and Sound Change. A Multimedia Reference Tool*. Berlin: Mouton de Gruyter.

Ladefoged, P. (1968). *A phonetic study of West African languages: An auditory-instrumental survey* (No. 1). Cambridge University Press.

Ladefoged, P. (2003). *A course in phonetics*. Fort Worth, London: Harcourt Brace.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245-294.

Lahiri, A., & Marslen-Wilson, W.D. (1992). Lexical processing and phonological representations. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 229-254). Cambridge: Cambridge University Press.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, *7*, 637-675.

Lapesa, R. (1981). *Historia de la lengua española. Novena edición corregida y aumentada*. Madrid: Gredos.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English1. *Journal of Sociolinguistics*, *15*(2), 256-268.

Lenz, R. (1940a [1892-1893]). Estudios Chilenos (Chilenische Studien) I – VII. In A. Alonso & R. Lida (Eds.), *El español en Chile. Trabajos de Rodolfo Lenz, Andrés Bello y Rodolfo Oroz* (pp. 85-208). Buenos Aires: Facultad de Filosofía y Letras de la Universidad de Buenos Aires.

Lenz, R. (1940b [1893]). Para el conocimiento del español de América (Beiträge zur Kenntnis des Amerikanospanisch). In A. Alonso & R. Lida (Eds.), *El español en Chile. Trabajos de Rodolfo Lenz, Andrés Bello y Rodolfo Oroz* (pp. 209-258). Buenos Aires: Facultad de Filosofía y Letras de la Universidad de Buenos Aires.

Leslau, W. (1969). Frequency as determinant of linguistic changes in the Ethiopian languages. *Word*, *25*(1-3), 180-189.

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and Information Systems*, *10*(4), 453-472.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172-187.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Lipski, J. M. (1984). On the weakening of /s/ in Latin American Spanish. *Zeitschrift für Dialektologie und Linguistik*, *51*(1), 31-43.

Lisker, L. (1986). “Voicing” in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 3-11.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America*, *49*, 606-08.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*(2), 874-886.

López Gavín, E. (2015). *Una revisión del sistema fonológico español: de Alarcos Llorach a la NGLE* (Unpublished doctoral thesis). Universidad de Santiago de Compostela, Galicia, Spain.

Lozano, M. D. (1978). *Stop and spirant alternations: Fortition and spirantization processes in phonology* (Unpublished doctoral thesis). Indiana University, Indiana, USA.

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, *49*(1), 11-49.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, *7*(03), 433-458.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide. Mahwah, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*(5), 452.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge university press.

Marslen-Wilson, W. D. (1973). Linguistic Structure and Speech Shadowing at Very Short Latencies. *Nature*, *244*, 522-523.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*(1-3), 55-73.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71-102.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29-63.

Martínez Celdrán, E. (1984). Cantidad e intensidad en los sonidos obstruyentes del castellano: hacia una caracterización acústica de los sonidos aproximantes. *Estudios de Fonética Experimental*, *1*, 71-129.

Martínez Celdrán, E. (1991). Sobre la naturaleza fonética de los alófonos de /b, d, g/ en español y sus distintas denominaciones. *Verba*, *18*, 235-253.

Martínez Celdrán, E. (2013). Caracterización acústica de las aproximantes espirantes en español. *Estudios de Fonética Experimental*, *22*, 11-36.

Martínez-Celdrán, E. (2004). Problems in the classification of approximants. *Journal of the International Phonetic Association*, *34*(02), 201-210.

Martínez-Celdrán, E., & Regueira, X. L. (2008). Spirant approximants in Galician. *Journal of the International Phonetic Association*, *38*(01), 51-68.

Mascaró, J. (1984). Continuant spreading in Basque, Catalan and Spanish. In M. Aronoff & R. Oehrle (Eds.), *Language Sound Structure: Studies Presented to Morris Halle by His Teacher and Students* (pp. 287-298). Cambridge, MA: MIT Press.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, *67*(3), 996-1013.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural “oddities” as artificial normality. *Perception & Psychophysics*, *70*(7), 1235-1242.

Mayo, C., & Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America, 115*, 3184-3194.

McClelland, J. L., & Elman, J. L. (1986a). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McClelland, J. L., & Elman, J. L. (1986b). Interactive processes in speech perception: The TRACE model. *Parallel Distributed Processing*, *2*(58), 121.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306.

McQueen, J. (1996). Phonetic categorisation. *Language and Cognitive Processes*, *11*(6), 655-664.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 433.

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The Handbook of Cognition* (pp. 255–275). London: Sage Publications.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113-1126.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*(1), 73-103.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Natsumeda, T., Mori, S., & Yuma, M. (2012). Size-mediated dominance and aggressive behavior of male Japanese fluvial sculpin Cottus pollux (Pisces: Cottidae) reduce nest-site abundance and mating success of conspecific rivals. *Journal of Ethology*, *30*(2), 239-245.

Nearey, T. M. (1977). *Phonetic Feature Systems for Vowels* (Unpublished doctoral thesis). University of Alberta, Edmonton, Alberta, Canada.

Newman, J. E., & Dell, G. S. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior*, *17*(3), 359-374.

Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, *67*(1), 276-287.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(03), 299-325.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191-243.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological review*, *85*(3), 172.

Ogden, R., & Local, J. K. (1994). Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics*, *30*(2), 477-498.

Ogura, T. (2012). Use of video system and its effects on abnormal behaviour in captive Japanese macaques (Macaca fuscata). *Applied Animal Behaviour Science*, *141*(3), 173-183.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189-216). Springer New York.

Oroz, R. (1966). *La lengua castellana en Chile*. Santiago: Editorial Universitaria.

Ortega-Llebaria, M. (2003). Effects of Phonetic and Inventory Constraints in the Spirantization of Intervocalic Voiced Stops: Comparing two Different Measurements of Energy Change. *Stress*, *7*(6), 5.

Parrell, B. (2010). Articulation from acoustics: Estimating constriction degree from the acoustic signal. *Journal of the Acoustical Society of America*, *128*(4), 2289.

Parrell, B. (2011). Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials. *Laboratory Phonology*, *2*(2), 423-449.

Parrell, B. (2012). How tongue posture differences affect reduction in coronals: Differences between Spanish and English. *Journal of the Acoustical Society of America*, *132*(3), 1936.

Pérez, E. H. (2003). Frecuencia de fonemas. *e-rthabla, Revista electrónica de Tecnología del Habla*, *1*. Retrieved from http://lorien.die.upm.es/~lapiz/e-rthabla/numeros/N1/N1_A4.pdf.

Pérez, H. E. (2007). Estudio de la variación estilística de la serie /b-d-g/ en posición intervocálica en el habla de los noticieros de la televisión chilena. *Estudios de Fonética Experimental*, *16*, 228-259.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*(2), 175-184.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*, *7*, 101-139.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Piñeros, C. E. (2002). Markedness and laziness in Spanish obstruents. *Lingua*, *112*(5), 379-413.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 1037.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 699.

Poniatowski, D., & Fartmann, T. (2009). Experimental evidence for density-determined wing dimorphism in two bush-crickets (Ensifera: Tettigoniidae). *European Journal of Entomology*, *106*(4), 599.

Poniatowski, D., & Fartmann, T. (2011). Weather-driven changes in population density determine wing dimorphism in a bush-cricket species. *Agriculture, Ecosystems & Environment*, *145*(1), 5-9.

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, *17*(2), 132-138.

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1), 103-121.

Quilis, A. (1999). *Tratado de fonología y fonética españolas*. Madrid: Gredos.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rabanales, A. (1992). El español de Chile: Situación actual. In C. Hernández Alonso (Ed.), *Historia y presente del español de América* (pp. 565-592). Valladolid: Junta de Castillo y León.

Rabanales, A. (2000). El español de Chile: presente y futuro. *Onomazein*, *5*, 135-141.

Radeau, M., & Morais, J. (1990). The uniqueness point effect in the shadowing of spoken words. *Speech Communication*, *9*(2), 155-164.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*(01), 55-97.

Real Academia Española (2014). *Banco de datos (CREA)* [online]. Corpus de referencia del español actual. <http://www.rae.es> [20 August 2015]

Repp, B. H. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *Journal of the Acoustical Society of America*, *74*, 420.

Romero Gallego, J. (1995). *Gestural organization in Spanish: An experimental study of spirantization and aspiration* (Unpublished doctoral thesis). University of Connecticut, Connecticut, USA.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147-1153.

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, *19*(5), 394-398.

Sadowsky, S. (2010). El alófono labiodental sonoro [v] del fonema /b/ en el castellano de Concepción (Chile): una investigación exploratoria. *Estudios de Fonética Experimental*, *19*, 231-261.

Sadowsky, S. (2015). Variación sociofonética de las consonantes del castellano chileno. *Sociolinguistic Studies*, *9*(1), 71-92.

Sadowsky, S., & Martínez Gamboa, R. (2004). *Lista de Frecuencias de Palabras del Castellano de Chile (Lifcach)*. Version 1.0/1.1. Retrieved November 19, 2012, from http://sadowsky.cl/lifcach.html

Sadowsky, S., & Salamanca, G. (2011). The phonetic inventory of Chilean Spanish: guiding principles, provisional inventory of consonants and system of representation (IPA-CL). *Onomazein*, *24*, 61-84.

Salas, A. (1996-1997). La lectura de noticias en la televisión chilena: modelo y norma en el fonetismo del castellano en Chile. *Anuario de Lingüística Hispánica*, *12-13*(2), 819-826.

Samuel, A. G. (1981a). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474.

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1124.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, *26*(1), 36-56.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes?. *Journal of Experimental Psychology: General*, *125*(1), 28.

Sánchez Lobato, J. (1994). El español en América. In J. Sánchez Lobato & I. Santos Gargallo (Eds.), *Problemas y métodos en la enseñanza del español como lengua extranjera*. Paper presented at the IV Congreso Internacional de la ASELE, Universidad Complutense de Madrid, Madrid, 7-9 October, 1993 (pp. 553-570), Centro Virtual Cervantes.

Santos, E. S., Maia, R., & Macedo, R. H. (2009). Condition-dependent resource value affects male–male competition in the blue–black grassquit. *Behavioral Ecology*, *20*(3), 553-559.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110-114.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1.

Schwab, J. A. (2002). *Multinomial logistic regression: Basic relationships and complete problems* [PowerPoint slides]. Retrieved from http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111-123.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, *20*(6), 615-627.

Silva-Fuenzalida, I. (1952-1953). Estudio fonológico del español de Chile. *Boletín de Filología*, *7*, 153-176.

Simonet, M., Hualde, J. I., & Nadeu, M. (2012, September). Lenition of /d/ in spontaneous Spanish and Catalan. In *INTERSPEECH* (pp. 1416-1419).

Sokal, R. R., & Rohlf, F. J. (2003). *Biometry, the principles and practice of statistics in biological research*, NY: W.H. Freeman and Company.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognitio*n, *14*(1), 17-26.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, *111*(4), 1872-1891.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1-38). Psychology Press.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219-238.

Swinney, D. A., & Hakes, D. T. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 681-689.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, *24*(3), 320-335.

Thünken, T., Meuthen, D., Bakker, T. C., & Kullmann, H. (2010). Parental investment in relation to offspring quality in the biparental cichlid fish Pelvicachromis taeniatus. *Animal Behaviour*, *80*(1), 69-74.

Titone, C. M. C. D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635-646.

Titze, I. R., & Winholtz, W. S. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech, Language, and Hearing Research*, *36*(6), 1177-1190.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2), 331-353.

Toscano, J. C., & McMurray, B. (2010). Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. *Cognitive Science*, *34*(3), 434-464.

Trisnawati, D. W., Tsukamoto, T., & Yasuda, H. (2015). Indirect effects of nutrients in organic and conventional paddy field soils on the rice grasshopper, Oxya japonica (Orthoptera: Acrididae), mediated by rice plant nutrients. *Applied Entomology and Zoology*, *50*(1), 99-107.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research*, *3*, 1–28.U

Valdivieso, H. (1993). Perfil fonético de escolares de Concepción. *Revista de Lingüística Teórica y Aplicada*, *31*, 119-135.

Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45-65.

Van Oers, K., Drent, P. J., Dingemanse, N. J., & Kempenaers, B. (2008). Personality is associated with extrapair paternity in great tits, Parus major. *Animal Behaviour*, *76*(3), 555-563.

Van Son, R. J., & Pols, L. C. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, *88*(4), 1683-1693.

Véliz, M. O., Araya, A., & Rodríguez, G. (1977). Muestra del español habado en las oficinas salitreras. *Estudios Filológicos*, *1*, 131-162.

Vennemann, T. (1988). *Preference laws for syllable structure: with special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.

Vergara Fernández, V. (2011). Producción de /v/ como alófono de /b/ en niños prealfabetizados de la provincia de Concepción (Chile). *Boletín de Lingüística*, 23(35-36), 123-142.

Vergara Fernández, V. (2013). Relación entre alfabetización y producción de los alófonos de /b/: estudio del habla cuidada de hablantes prealfabetizados y alfabetizados. *Onomázein*, *27*, 158-170.

Vergara, V., & Pérez, H. E. (2013). Estudio de la incidencia de la representación gráfica (escritura) en la producción del alófono labiodental [v] del fonema /b/. *Boletín de Filología*, *48*(2), 119-128.

Vivar, P., & León Valdés, H. (2009). Desarrollo fonológico-fonético en un grupo de niños entre 3 y 5, 11 años. *Revista CEFAC. 11*(2), 190-8.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.

Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, *16*(1), 150-156.

Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, *92*(1), 3-10.

Weenink, D. (2009). The KlattGrid speech synthesizer. In *Interspeech*, *10*, 2059-2062. International Speech Communication Association.

Werker, J. F., & Polka, L. (1993). The ontogeny and developmental significance of language-specific phonetic perception. In *Developmental neurocognition: speech and face processing in the first year of life* (pp. 275-288). Springer Netherlands.

Wigdorsky, L. (1978). Realización de algunos fonemas consonánticos en el castellano de Santiago: informe preliminar. *Revista de Lingüística Teórica y Aplicada*, *16*, 53-60.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289-297.

Yeni–Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, *70*(4), 966-975.

Yuan, J., Liberman, M., & Cieri, C. (2006, September). *Towards an integrated understanding of speaking rate in conversation*. Paper presented at the International Conference on Spoken Language Processing (Interspeech), Pittsburgh.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America*, *94*(4), 1966-1982.

Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, 470-481.

Zampini, M. L. (1998). The Relationship between the Production and Perception of L2 Spanish Stops. *Texas Papers in Foreign Language Education*, *3*(3), 85-100.

Zwicky, A. (1972). Note on a Phonological Hierarchy in English. In R. Stockwell & R. Macaulay (Eds) *Linguistic change and generative theory*. Indiana University Press.

Appendixes

Appendix 1: Full list of continua from Chapter 6

Table A1.1. Full specifications of all continua used in perception tasks, Chapter 6. Table contents: unique continuum identification number; task ("PM": phoneme monitoring, "ID": identification; "DISC": discrimination); experimental condition; status of the task (practice or experimental); reference word for the full approximant; prime for the approximant consonant; stimuli for the full approximant; reference word for the elided variant; prime for the elided variant; stimuli for the elided variant; total number of stimulus steps or stimulus pairs for given continuum; repetitions per step or pair; and format.

N.	Task	Phon.	Condition	Status	Full word (reference)	Prime full	Stimuli full	Elided word (reference)	Prime elided	Stimuli elided	Steps / pairs	Reps.	Format
1	PM	/b/	Segmental	Practice	*releva*		[ˈe.β̞a]	*relea*		[ˈe.a]	10	2	
2	PM	/b/	Segmental	Task	*cubetazo*		[ˈu.β̞e]	*cuetazo*		[we]	10	2	
3	PM	/b/	Word-level	Practice	*releva*		[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	10	2	
4	PM	/b/	Word-level	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	10	2	
5	PM	/b/	Primed app.	Practice	*releva*	*reemplazar*	[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	10	2	
6	PM	/b/	Primed app.	Task	*cubetazo*	*balde*	[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	10	2	
7	PM	/b/	Primed elided	Practice	*releva*		[re.ˈle.β̞a]	*relea*	*repasar*	[re.ˈle.a]	10	2	
8	PM	/b/	Primed elided	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*	*explosión*	[kwe.ˈta.so]	10	2	
9	PM	/d/	Segmental	Practice	*callado*		[ˈa.ð̞o]	*Callao*		[ˈa.o]	10	2	
10	PM	/d/	Segmental	Task	*dudo*		[ˈu.ð̞o]	*dúo*		[ˈu.o]	10	2	

N.	Task	Phon.	Condition	Status	Full word (reference)	Prime full	Stimuli full	Elided word (reference)	Prime elided	Stimuli elided	Steps / pairs	Reps.	Format
11	PM	/d/	Word-level	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	10	2	
12	PM	/d/	Word-level	Task	*dudo*		[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	10	2	
13	PM	/d/	Primed app.	Practice	*callado*	*enmudecer*	[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	10	2	
14	PM	/d/	Primed app.	Task	*dudo*	*titubear*	[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	10	2	
15	PM	/d/	Primed elided	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*	*puerto*	[ka.ˈʝ̞a.o]	10	2	
16	PM	/d/	Primed elided	Task	*dudo*		[ˈdu.ð̞o]	*dúo*	*pareja*	[ˈdu.o]	10	2	
17	PM	/g/	Segmental	Practice	*mega*		[ˈe.ɣ̞a]	*mea*		[ˈe.a]	10	2	
18	PM	/g/	Segmental	Task	*boga*		[ˈo.ɣ̞a]	*boa*		[ˈo.a]	10	2	
19	PM	/g/	Word-level	Practice	*mega*		[ˈme.ɣ̞a]	*mea*		[ˈme.a]	10	2	
20	PM	/g/	Word-level	Task	*boga*		[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	10	2	
21	PM	/g/	Primed app.	Practice	*mega*	*grande*	[ˈme.ɣ̞a]	*mea*		[ˈme.a]	10	2	
22	PM	/g/	Primed app.	Task	*boga*	*actualidad*	[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	10	2	
23	PM	/g/	Primed elided	Practice	*mega*		[ˈme.ɣ̞a]	*mea*	*orinar*	[ˈme.a]	10	2	
24	PM	/g/	Primed elided	Task	*boga*		[ˈbo.ɣ̞a]	*boa*	*constrictor*	[ˈbo.a]	10	2	
25	ID	/b/	Segmental	Practice	*releva*		[ˈe.β̞a]	*relea*		[ˈe.a]	10	2	
26	ID	/b/	Segmental	Task	*cubetazo*		[ˈu.β̞e]	*cuetazo*		[we]	10	2	
27	ID	/b/	Word-level	Practice	*releva*		[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	10	2	
28	ID	/b/	Word-level	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	10	2	
29	ID	/b/	Primed app.	Practice	*releva*	*reemplazar*	[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	10	2	
30	ID	/b/	Primed app.	Task	*cubetazo*	*balde*	[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	10	2	
31	ID	/b/	Primed elided	Practice	*releva*		[re.ˈle.β̞a]	*relea*	*repasar*	[re.ˈle.a]	10	2	
32	ID	/b/	Primed elided	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*	*explosión*	[kwe.ˈta.so]	10	2	
33	ID	/d/	Segmental	Practice	*callado*		[ˈa.ð̞o]	*Callao*		[ˈa.o]	10	2	
34	ID	/d/	Segmental	Task	*dudo*		[ˈu.ð̞o]	*dúo*		[ˈu.o]	10	2	

N.	Task	Phon.	Condition	Status	Full word (reference)	Prime full	Stimuli full	Elided word (reference)	Prime elided	Stimuli elided	Steps / pairs	Reps.	Format
35	ID	/d/	Word-level	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	10	2	
36	ID	/d/	Word-level	Task	*dudo*		[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	10	2	
37	ID	/d/	Primed app.	Practice	*callado*	*enmudecer*	[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	10	2	
38	ID	/d/	Primed app.	Task	*dudo*	*titubear*	[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	10	2	
39	ID	/d/	Primed elided	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*	*puerto*	[ka.ˈʝ̞a.o]	10	2	
40	ID	/d/	Primed elided	Task	*dudo*		[ˈdu.ð̞o]	*dúo*	*pareja*	[ˈdu.o]	10	2	
41	ID	/g/	Segmental	Practice	*mega*		[ˈe.ɣ̞a]	*mea*		[ˈe.a]	10	2	
42	ID	/g/	Segmental	Task	*boga*		[ˈo.ɣ̞a]	*boa*		[ˈo.a]	10	2	
43	ID	/g/	Word-level	Practice	*mega*		[ˈme.ɣ̞a]	*mea*		[ˈme.a]	10	2	
44	ID	/g/	Word-level	Task	*boga*		[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	10	2	
45	ID	/g/	Primed app.	Practice	*mega*	*grande*	[ˈme.ɣ̞a]	*mea*		[ˈme.a]	10	2	
46	ID	/g/	Primed app.	Task	*boga*	*actualidad*	[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	10	2	
47	ID	/g/	Primed elided	Practice	*mega*		[ˈme.ɣ̞a]	*mea*	*orinar*	[ˈme.a]	10	2	
48	ID	/g/	Primed elided	Task	*boga*		[ˈbo.ɣ̞a]	*boa*	*constrictor*	[ˈbo.a]	10	2	
49	DISC	/b/	Segmental	Practice	*releva*		[ˈe.β̞a]	*relea*		[ˈe.a]	7	4	ABX
50	DISC	/b/	Segmental	Task	*cubetazo*		[ˈu.β̞e]	*cuetazo*		[we]	7	4	ABX
51	DISC	/b/	Word-level	Practice	*releva*		[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	7	4	ABX
52	DISC	/b/	Word-level	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	7	4	ABX
53	DISC	/b/	Primed app.	Practice	*releva*	*reemplazar*	[re.ˈle.β̞a]	*relea*		[re.ˈle.a]	7	2	ABX
54	DISC	/b/	Primed app.	Task	*cubetazo*	*balde*	[ku.β̞e.ˈta.so]	*cuetazo*		[kwe.ˈta.so]	7	2	ABX
55	DISC	/b/	Primed elided	Practice	*releva*		[re.ˈle.β̞a]	*relea*	*repasar*	[re.ˈle.a]	7	2	ABX
56	DISC	/b/	Primed elided	Task	*cubetazo*		[ku.β̞e.ˈta.so]	*cuetazo*	*explosión*	[kwe.ˈta.so]	7	2	ABX
57	DISC	/d/	Segmental	Practice	*callado*		[ˈa.ð̞o]	*Callao*		[ˈa.o]	7	4	ABX
58	DISC	/d/	Segmental	Task	*dudo*		[ˈu.ð̞o]	*dúo*		[ˈu.o]	7	4	ABX

N.	Task	Phon.	Condition	Status	Full word (reference)	Prime full	Stimuli full	Elided word (reference)	Prime elided	Stimuli elided	Steps / pairs	Reps.	Format
59	DISC	/d/	Word-level	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	7	4	ABX
60	DISC	/d/	Word-level	Task	*dudo*		[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	7	4	ABX
61	DISC	/d/	Primed app.	Practice	*callado*	*enmudecer*	[ka.ˈʝ̞a.ð̞o]	*Callao*		[ka.ˈʝ̞a.o]	7	2	ABX
62	DISC	/d/	Primed app.	Task	*dudo*	*titubear*	[ˈdu.ð̞o]	*dúo*		[ˈdu.o]	7	2	ABX
63	DISC	/d/	Primed elided	Practice	*callado*		[ka.ˈʝ̞a.ð̞o]	*Callao*	*puerto*	[ka.ˈʝ̞a.o]	7	2	ABX
64	DISC	/d/	Primed elided	Task	*dudo*		[ˈdu.ð̞o]	*dúo*	*pareja*	[ˈdu.o]	7	2	ABX
65	DISC	/g/	Segmental	Practice	*mega*		[ˈe.ɣ̞a]	*mea*		[ˈe.a]	7	4	ABX
66	DISC	/g/	Segmental	Task	*boga*		[ˈo.ɣ̞a]	*boa*		[ˈo.a]	7	4	ABX
67	DISC	/g/	Word-level	Practice	*mega*		[ˈme.ɣ̞a]	*mea*		[ˈme.a]	7	4	ABX
68	DISC	/g/	Word-level	Task	*boga*		[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	7	4	ABX
69	DISC	/g/	Primed app.	Practice	*mega*	*grande*	[ˈme.ɣ̞a]	*mea*		[ˈme.a]	7	2	ABX
70	DISC	/g/	Primed app.	Task	*boga*	*actualidad*	[ˈbo.ɣ̞a]	*boa*		[ˈbo.a]	7	2	ABX
71	DISC	/g/	Primed elided	Practice	*mega*		[ˈme.ɣ̞a]	*mea*	*orinar*	[ˈme.a]	7	2	ABX
72	DISC	/g/	Primed elided	Task	*boga*		[ˈbo.ɣ̞a]	*boa*	*constrictor*	[ˈbo.a]	7	2	ABX